GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 8, 2004, 11:00:44 ; Search time 28.7778 Seconds
(without alignments)
76.748 Million cell updates/sec

Title: US-09-753-139C-1
Perfect score: 42
Sequence: 1 CXCXPHP 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	95.2	189	4	Q9HD73	Q9hd73 homo sapien
2	40	95.2	445	4	Q96PY0	Q96py0 homo sapien
3	38	90.5	225	12	Q9IGT4	Q9igt4 porcine ade
4	38	90.5	527	5	P91773	P91773 penaeus jap
5	37	88.1	110	11	Q8BS77	Q8bs77 mus musculu
6	34	81.0	289	10	Q7XM15	Q7xm15 oryza sativ
7	33	78.6	190	11	Q8BJA8	Q8bja8 mus musculu
8	33	78.6	246	5	Q17792	Q17792 caenorhabdi
9	33	78.6	1376	12	Q83330	Q83330 murine hepa
10	33	78.6	1427	11	Q8VIB7	Q8vib7 mesocricetu
11	32	76.2	40	7	Q860N3	Q860n3 rattus sp.
12	32	76.2	50	9	Q854K7	Q854k7 mycobacteri
13	32	76.2	89	10	P82627	P82627 arabidopsis
14	32	76.2	101	5	Q8IMN0	Q8imn0 drosophila
15	32	76.2	103	17	Q9YG80	Q9yg80 aeropyrum p
16	32	76.2	119	5	O76648	O76648 caenorhabdi
17	32	76.2	134	4	Q8N788	Q8n788 homo sapien
18	32	76.2	193	4	Q86TC4	Q86tc4 homo sapien
19	32	76.2	194	4	Q9BV29	Q9bv29 homo sapien
20	32	76.2	194	4	Q8NAR7	Q8nar7 homo sapien
21	32	76.2	246	10	Q8RYG1	Q8ryg1 oryza sativ
22	32	76.2	304	2	Q53405	Q53405 acetobacter
23	32	76.2	311	5	Q9W2C2	Q9w2c2 drosophila
24	32	76.2	335	12	Q8JKG9	Q8jkg9 heliothis z
25	32	76.2	337	4	Q8NHD3	Q8nhd3 homo sapien
26	32	76.2	342	4	Q8NHD5	Q8nhd5 homo sapien
27	32	76.2	344	2	Q841K9	Q841k9 streptomyce
28	32	76.2	351	4	Q8TF26	Q8tf26 homo sapien
29	32	76.2	353	16	Q8Y3G4	Q8y3g4 ralstonia s
30	32	76.2	362	16	Q7V5J8	Q7v5j8 prochloroco
31	32	76.2	401	4	Q9H1U4	Q9h1u4 homo sapien
32	32	76.2	434	10	Q8S1U4	Q8s1u4 oryza sativ
33	32	76.2	442	10	Q8S1U6	Q8s1u6 oryza sativ
34	32	76.2	443	10	Q8S1U7	Q8s1u7 oryza sativ
35	32	76.2	461	5	Q9VZN1	Q9vzn1 drosophila
36	32	76.2	569	4	Q8NHD4	Q8nhd4 homo sapien
37	32	76.2	581	5	Q9BL07	Q9bl07 caenorhabdi
38	32	76.2	582	10	Q9AX44	Q9ax44 oryza sativ
39	32	76.2	600	11	Q8BH27	Q8bh27 mus musculu
40	32	76.2	610	16	Q83QL8	Q83ql8 shigella fl
41	32	76.2	647	5	Q86NW5	Q86nw5 drosophila
42	32	76.2	744	4	Q8NHD2	Q8nhd2 homo sapien
43	32	76.2	866	4	Q8IXF3	Q8ixf3 homo sapien
44	32	76.2	1275	11	Q99PW0	Q99pw0 rattus norv
45	32	76.2	1432	11	Q99J86	Q99j86 rattus norv

ALIGNMENTS

RESULT 1
Q9HD73
ID Q9HD73 PRELIMINARY; PRT; 189 AA.
AC Q9HD73;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Ornithine decarboxylase antizyme.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang W., Wan T., Cao X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF242521; AAF99601.1; -.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0008073; F:ornithine decarboxylase inhibitor activity; IEA.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR002993; ODC_AZ.
DR Pfam; PF02100; ODC_AZ; 1.
DR ProDom; PD007483; ODC_AZ; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS01337; ODC_AZ; 1.
SQ SEQUENCE 189 AA; 20999 MW; 232909D6AD01FEEF CRC64;

Query Match 95.2%; Score 40; DB 4; Length 189;
Best Local Similarity 71.4%; Pred. No. 9.8;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPHP 7
              | | |||
Db         31 CSCAPHP 37
```

RESULT 2
Q96PY0
ID Q96PY0 PRELIMINARY; PRT; 445 AA.

```
AC   Q96PY0;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein KIAA1908 (Fragment).
GN   KIAA1908.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21456161; PubMed=11572484;
RA   Nagase T., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XXI.
RT   The complete sequences of 60 new cDNA clones from brain which code for
RT   large proteins.";
RL   DNA Res. 8:179-187(2001).
DR   EMBL; AB067495; BAB67801.1; -.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   445 AA;  47372 MW;  FE0DF053B19AF891 CRC64;

  Query Match             95.2%;  Score 40;  DB 4;  Length 445;
  Best Local Similarity   71.4%;  Pred. No. 19;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         48 CSCSPHP 54


RESULT 3
Q9IGT4
ID   Q9IGT4      PRELIMINARY;      PRT;   225 AA.
AC   Q9IGT4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   IVa2 (Fragment).
GN   IVA2.
OS   Porcine adenovirus type 3 (PAV-3).
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=35265;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Aggarwal N., Mittal S.K.;
RT   "Sequence Analysis of Porcine Adenovirus Type 3 E1 Region, pIX, pIVa2
RT   Genes, and Five Novel Open Reading Frames.";
RL   Intervirology 0:0-0(2000).
DR   EMBL; AF247039; AAF78234.1; -.
DR   GO; GO:0019083; P:viral transcription; IEA.
DR   InterPro; IPR003389; Adeno_IVa2.
DR   Pfam; PF02456; Adeno_IVa2; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   225 AA;  26243 MW;  112CDEB2E007F3B9 CRC64;

  Query Match             90.5%;  Score 38;  DB 12;  Length 225;
  Best Local Similarity   71.4%;  Pred. No. 25;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db        169 CTCPPHP 175


RESULT 4
P91773
ID   P91773      PRELIMINARY;      PRT;   527 AA.
AC   P91773;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Pjchi-2.
OS   Penaeus japonicus (Kuruma prawn).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC   Penaeidae; Marsupenaeus.
OX   NCBI_TaxID=27405;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Watanabe T., Kono M.;
RT   "Isolation of a cDNA Encoding a Chitinase Family Protein from
RT   Cuticular Tissues of the Kuruma Prawn Penaeus japonicus.";
RL   Zool. Sci. 0:0-0(1996).
DR   EMBL; D89751; BAA14014.1; -.
DR   GO; GO:0005576; C:extracellular; IEA.
DR   GO; GO:0008061; F:chitin binding; IEA.
DR   GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR   GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR   GO; GO:0006030; P:chitin metabolism; IEA.
DR   InterPro; IPR002557; Chitin_bind_PerA.
DR   InterPro; IPR001223; Glyco_hydro_18.
DR   InterPro; IPR001579; Glyco_hydro_18AS.
DR   Pfam; PF01607; CBM_14; 1.
DR   Pfam; PF00704; Glyco_hydro_18; 1.
DR   ProDom; PD000471; Glyco_hydro_18; 1.
DR   SMART; SM00494; ChtBD2; 1.
DR   SMART; SM00636; Glyco_18; 1.
DR   PROSITE; PS01095; CHITINASE_18; 1.
KW   Glycosidase; Hydrolase.
SQ   SEQUENCE   527 AA;  59162 MW;  B9CBAEAB8CDF8710 CRC64;

  Query Match             90.5%;  Score 38;  DB 5;  Length 527;
  Best Local Similarity   71.4%;  Pred. No. 49;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db        500 CGCEPHP 506


RESULT 5
Q8BS77
ID   Q8BS77      PRELIMINARY;      PRT;   110 AA.
AC   Q8BS77;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK035013; BAC28910.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   110 AA;  12073 MW;  B956CDC7C1FC894C CRC64;

  Query Match             88.1%;  Score 37;  DB 11;  Length 110;
  Best Local Similarity   71.4%;  Pred. No. 21;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         79 CWCQPHP 85
```

```
RESULT 6
Q7XM15
ID   Q7XM15      PRELIMINARY;      PRT;   289 AA.
AC   Q7XM15;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   OSJNBa0084K01.16 protein.
GN   OSJNBA0084K01.16.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA   Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA   Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA   Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA   Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA   Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA   Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA   Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA   Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL606999; CAE04844.1; -.
SQ   SEQUENCE   289 AA;  31003 MW;  2E52E631ACB445C3 CRC64;

  Query Match             81.0%;  Score 34;  DB 10;  Length 289;
  Best Local Similarity   57.1%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db        279 CSCSPYP 285


RESULT 7
Q8BJA8
ID   Q8BJA8      PRELIMINARY;      PRT;   190 AA.
AC   Q8BJA8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical ankyrin repeat region circular profile/yeast DNA-binding
DE   domain containing protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NOD; TISSUE=Spleen;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK089667; BAC40941.1; -.
DR   InterPro; IPR002110; ANK.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   190 AA;  19996 MW;  2D11E31D709E38FA CRC64;

  Query Match             78.6%;  Score 33;  DB 11;  Length 190;
  Best Local Similarity   66.7%;  Pred. No. 1.5e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        163 CACAPH 168


RESULT 8
Q17792
ID   Q17792      PRELIMINARY;      PRT;   246 AA.
AC   Q17792;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein C07E3.6.
GN   C07E3.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matthews P.;
RL   Submitted (JUN-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z49908; CAA90100.2; -.
DR   PIR; T19056; T19056.
DR   WormPep; C07E3.6; CE32306.
KW   Hypothetical protein.
SQ   SEQUENCE   246 AA;  26490 MW;  D7EB56BA33F534A9 CRC64;

  Query Match             78.6%;  Score 33;  DB 5;  Length 246;
  Best Local Similarity   66.7%;  Pred. No. 1.9e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         49 CSCTPH 54


RESULT 9
Q83330
ID   Q83330      PRELIMINARY;      PRT;  1376 AA.
AC   Q83330;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   S glycoprotein.
OS   Murine hepatitis virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11138;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RI;
RX   MEDLINE=95304829; PubMed=7785316;
RA   Kunita S., Zhang L., Homberger F.R., Compton S.R.;
RT   "Molecular characterization of the S proteins of two enterotropic
RT   murine coronavirus strains.";
RL   Virus Res. 35:277-289(1995).
DR   EMBL; U14645; AAA87062.1; -.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01601; Corona_S2; 1.
SQ   SEQUENCE   1376 AA;  151615 MW;  E9EE02151DBD7D9D CRC64;

  Query Match             78.6%;  Score 33;  DB 12;  Length 1376;
  Best Local Similarity   57.1%;  Pred. No. 7.5e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
```

```
Db        532 CTCAPNP 538


RESULT 10
Q8VIB7
ID   Q8VIB7      PRELIMINARY;      PRT;  1427 AA.
AC   Q8VIB7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Attractin.
GN   ATRN.
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21635551; PubMed=11773967;
RA   Kuramoto T., Nomoto T., Fujiwara A., Mizutani M., Sugimura T.,
RA   Ushijima T.;
RT   "Insertional mutation of the Attractin gene in the black tremor
RT   hamster.";
RL   Mamm. Genome 13:36-40(2002).
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
DR   EMBL; AB062913; BAB72012.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   GO; GO:0005198; F:structural molecule activity; IEA.
DR   GO; GO:0005529; F:sugar binding; IEA.
DR   GO; GO:0007275; P:development; IEA.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006652; Kelch_rep.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   Pfam; PF00431; CUB; 1.
DR   Pfam; PF01344; Kelch; 6.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF01437; PSI; 4.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00042; CUB; 1.
DR   SMART; SM00423; PSI; 5.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   Laminin EGF-like domain.
SQ   SEQUENCE   1427 AA;  158024 MW;  31F865993BA17B68 CRC64;

  Query Match             78.6%;  Score 33;  DB 11;  Length 1427;
  Best Local Similarity   66.7%;  Pred. No. 7.7e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        305 CSCSPH 310


RESULT 11
Q860N3
ID   Q860N3      PRELIMINARY;      PRT;    40 AA.
AC   Q860N3;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Major histocompatibility complex class I Eu-like protein
DE   (Fragment).
OS   Rattus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10118;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96217551; PubMed=8613150;
RA   Yuan X.J., Salgar S.K., Hassett A.L., McHugh K.P., Kunz H.W.,
RA   Gill T.J. III.;
RT   "Physical mapping of the E/C and grc regions of the rat major
RT   histocompatibility complex.";
RL   Immunogenetics 44:9-18(1996).
DR   EMBL; S81844; AAP32239.1; -.
FT   NON_TER       1      1
FT   NON_TER      40     40
SQ   SEQUENCE   40 AA;  4957 MW;  FF31A75A18D217E3 CRC64;

  Query Match             76.2%;  Score 32;  DB 7;  Length 40;
  Best Local Similarity   66.7%;  Pred. No. 65;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         10 CTCGPH 15


RESULT 12
Q854K7
ID   Q854K7      PRELIMINARY;      PRT;    50 AA.
AC   Q854K7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Gp57.
OS   Mycobacteriophage Omega.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=205879;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22592660; PubMed=12705866;
RA   Pedulla M.L., Ford M.E., Houtz J.M., Karthikeyan T., Wadsworth C.,
RA   Lewis J.A., Jacobs-Sera D., Falbo J., Gross J., Pannunzio N.R.,
RA   Brucker W., Kumar V., Kandasamy J., Keenan L., Bardarov S.,
RA   Kriakov J., Lawrence J.G., Jacobs W.R. Jr., Hendrix R.W.,
RA   Hatfull G.F.;
RT   "Origins of highly mosaic mycobacteriophage genomes.";
RL   Cell 113:171-182(2003).
DR   EMBL; AY129338; AAN12701.1; -.
DR   GO; GO:0004872; F:receptor activity; IEA.
DR   InterPro; IPR001368; TNFR_c6.
DR   PROSITE; PS00652; TNFR_NGFR_1; 1.
SQ   SEQUENCE   50 AA;  5273 MW;  F5D9305CDF41EF00 CRC64;

  Query Match             76.2%;  Score 32;  DB 9;  Length 50;
  Best Local Similarity   66.7%;  Pred. No. 78;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         13 CTCFPH 18


RESULT 13
P82627
ID   P82627      PRELIMINARY;      PRT;    89 AA.
AC   P82627;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCRL8 precursor.
GN   SCRL8.
OS   Arabidopsis thaliana (Mouse-ear cress).
```

```
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016719; PubMed=11130712;
RA   Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA   White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA   Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA   Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA   Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA   Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA   Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA   Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA   Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA   Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA   Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA   Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA   Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA   Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA   Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA   Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT   "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:816-820(2000).
RN   [2]
RP   IDENTIFICATION.
RX   PubMed=11437247;
RA   Vanoosthuyse V., Miege C., Dumas C., Cock J.M.;
RT   "Two large Arabidopsis thaliana gene families are homologous to the
RT   Brassica gene superfamily that encodes pollen coat proteins and the
RT   male component of the self-incompatibility response.";
RL   Plant Mol. Biol. 46:17-34(2001).
DR   EMBL; AC018908; -; NOT_ANNOTATED_CDS.
KW   Hypothetical protein; Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   CHAIN        24     89       HYPOTHETICAL PROTEIN SCRL8.
SQ   SEQUENCE   89 AA;  10183 MW;  E7B19B65AEAB497F CRC64;

  Query Match             76.2%;  Score 32;  DB 10;  Length 89;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+02;
  Matches    4;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | | |
Db         67 CTCTPQP 73


RESULT 14
Q8IMN0
ID   Q8IMN0      PRELIMINARY;      PRT;   101 AA.
AC   Q8IMN0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG31308-PA.
GN   CG31308.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003762; AAN14130.1; -.
DR   FlyBase; FBgn0051308; CG31308.
SQ   SEQUENCE   101 AA;  11247 MW;  766D30410CBFA209 CRC64;

  Query Match             76.2%;  Score 32;  DB 5;  Length 101;
```

```
  Best Local Similarity   66.7%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         77 CVCAPH 82


RESULT 15
Q9YG80
ID   Q9YG80      PRELIMINARY;      PRT;   103 AA.
AC   Q9YG80;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein APE0021.
GN   APE0021.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC   Desulfurococcaceae; Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
DR   EMBL; AP000058; BAA78930.1; -.
DR   PIR; H72753; H72753.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   103 AA;  10953 MW;  84D9605698B6A996 CRC64;

  Query Match             76.2%;  Score 32;  DB 17;  Length 103;
  Best Local Similarity   57.1%;  Pred. No. 1.4e+02;
  Matches    4;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | |  ||
Db         13 CSCTTHP 19


Search completed: April  8, 2004, 11:16:13
Job time : 29.7778 secs
```

OM protein - protein search, using sw model

Run on: April 8, 2004, 11:16:30 ; Search time 30.3333 Seconds
(without alignments)
60.678 Million cell updates/sec

Title: US-09-753-139C-1
Perfect score: 42
Sequence: 1 CXCXPHP 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1073127 seqs, 262937947 residues

Total number of hits satisfying chosen parameters: 1073127

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	39	92.9	30	14	US-10-011-585A-157	Sequence 157, App
2	39	92.9	89	12	US-10-424-599-254963	Sequence 254963,
3	39	92.9	519	15	US-10-108-260A-3767	Sequence 3767, Ap
4	38	90.5	7	10	US-09-753-139C-1	Sequence 1, Appli
5	37	88.1	170	15	US-10-264-237-1473	Sequence 1473, Ap
6	34	81.0	289	12	US-10-412-699B-1681	Sequence 1681, Ap
7	34	81.0	289	15	US-10-374-780A-1670	Sequence 1670, Ap
8	33	78.6	91	12	US-10-424-599-153612	Sequence 153612,
9	33	78.6	97	9	US-09-925-297-595	Sequence 595, App
10	33	78.6	108	9	US-09-796-692-1071	Sequence 1071, Ap
11	33	78.6	108	14	US-10-040-862-1071	Sequence 1071, Ap
12	33	78.6	108	15	US-10-057-475B-1071	Sequence 1071, Ap
13	33	78.6	108	15	US-10-154-884B-1071	Sequence 1071, Ap
14	33	78.6	119	9	US-09-796-692-1736	Sequence 1736, Ap
15	33	78.6	119	14	US-10-040-862-1736	Sequence 1736, Ap
16	33	78.6	119	15	US-10-057-475B-1736	Sequence 1736, Ap
17	33	78.6	119	15	US-10-154-884B-1736	Sequence 1736, Ap
18	33	78.6	120	9	US-09-796-692-1065	Sequence 1065, Ap
19	33	78.6	120	14	US-10-040-862-1065	Sequence 1065, Ap
20	33	78.6	120	15	US-10-057-475B-1065	Sequence 1065, Ap
21	33	78.6	120	15	US-10-154-884B-1065	Sequence 1065, Ap
22	33	78.6	139	9	US-09-796-692-755	Sequence 755, App
23	33	78.6	139	14	US-10-040-862-755	Sequence 755, App
24	33	78.6	139	15	US-10-057-475B-755	Sequence 755, App
25	33	78.6	139	15	US-10-154-884B-755	Sequence 755, App
26	33	78.6	149	9	US-09-796-692-1077	Sequence 1077, Ap
27	33	78.6	149	14	US-10-040-862-1077	Sequence 1077, Ap
28	33	78.6	149	15	US-10-057-475B-1077	Sequence 1077, Ap
29	33	78.6	149	15	US-10-154-884B-1077	Sequence 1077, Ap
30	33	78.6	155	9	US-09-796-692-1642	Sequence 1642, Ap
31	33	78.6	155	14	US-10-040-862-1642	Sequence 1642, Ap
32	33	78.6	155	15	US-10-057-475B-1642	Sequence 1642, Ap
33	33	78.6	155	15	US-10-154-884B-1642	Sequence 1642, Ap
34	33	78.6	592	14	US-10-288-556-18	Sequence 18, Appl
35	33	78.6	665	12	US-10-424-599-149995	Sequence 149995,
36	32	76.2	40	10	US-09-820-649-143	Sequence 143, App
37	32	76.2	40	14	US-10-160-162-143	Sequence 143, App
38	32	76.2	57	12	US-10-424-599-262948	Sequence 262948,
39	32	76.2	114	12	US-10-424-599-223258	Sequence 223258,
40	32	76.2	133	12	US-10-424-599-211053	Sequence 211053,
41	32	76.2	135	13	US-10-001-843-158	Sequence 158, App
42	32	76.2	149	9	US-09-925-297-538	Sequence 538, App
43	32	76.2	164	14	US-10-106-698-5783	Sequence 5783, Ap
44	32	76.2	172	14	US-10-029-386-33998	Sequence 33998, A
45	32	76.2	173	14	US-10-029-386-34054	Sequence 34054, A

ALIGNMENTS

RESULT 1
US-10-011-585A-157
; Sequence 157, Application US/10011585A
; Publication No. US20030039986A1
; GENERAL INFORMATION:
; APPLICANT: Sun, Yongming
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific
; TITLE OF INVENTION: Genes and Proteins
; FILE REFERENCE: DEX-0261
; CURRENT APPLICATION NUMBER: US/10/011,585A
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/245,740
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 245
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 157
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-011-585A-157

Query Match 92.9%; Score 39; DB 14; Length 30;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPHP 7
              | | |||
Db         24 CACFPHP 30
```

RESULT 2
US-10-424-599-254963
; Sequence 254963, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:

```
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 254963
;   LENGTH: 89
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_72254C.1.pep
US-10-424-599-254963

  Query Match             92.9%;  Score 39;  DB 12;  Length 89;
  Best Local Similarity   71.4%;  Pred. No. 58;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         53 CTCRPHP 59


RESULT 3
US-10-108-260A-3767
; Sequence 3767, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3767
;   LENGTH: 519
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-108-260A-3767

  Query Match             92.9%;  Score 39;  DB 15;  Length 519;
  Best Local Similarity   71.4%;  Pred. No. 2.3e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         34 CSCLPHP 40


RESULT 4
US-09-753-139C-1
; Sequence 1, Application US/09753139C
; Publication No. US20030073808A1
; GENERAL INFORMATION:
;  APPLICANT: Quirk, Stephen
;  APPLICANT:  Tyrrell, David
;  TITLE OF INVENTION: Design and Use of Advanced Zinc Chelating Peptides to Regulate Ma
;  TITLE OF INVENTION:  Metalloproteinases
;  FILE REFERENCE: 44039-227522    11301-0200
;  CURRENT APPLICATION NUMBER: US/09/753,139C
;  CURRENT FILING DATE:  2000-12-29
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 7
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic Peptide
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (2)..(2)
;   OTHER INFORMATION: X = Ser or Thr
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (4)..(4)
;   OTHER INFORMATION: X = Ser, Ala or Val
US-09-753-139C-1

  Query Match             90.5%;  Score 38;  DB 10;  Length 7;
  Best Local Similarity   100.0%;  Pred. No. 9.7e+05;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              |||||||
Db          1 CXCXPHP 7


RESULT 5
US-10-264-237-1473
; Sequence 1473, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
;  APPLICANT: Birse et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA131P1
;  CURRENT APPLICATION NUMBER: US/10/264,237
;  CURRENT FILING DATE:  2002-10-04
;  PRIOR APPLICATION NUMBER: PCT/US01/16450
;  PRIOR FILING DATE: 2001-05-18
;  PRIOR APPLICATION NUMBER: US 60/205,515
;  PRIOR FILING DATE: 2000-05-19
;  NUMBER OF SEQ ID NOS: 2876
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 1473
;   LENGTH: 170
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (46)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
;   FEATURE:
;   NAME/KEY: MISC_FEATURE
;   LOCATION: (126)
;   OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-237-1473

  Query Match             88.1%;  Score 37;  DB 15;  Length 170;
  Best Local Similarity   71.4%;  Pred. No. 1.9e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db        114 CLCPPHP 120


RESULT 6
US-10-412-699B-1681
; Sequence 1681, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
;  APPLICANT: Mendel Biotechnology, Inc.
;  APPLICANT:  Zhang, James
;  APPLICANT:  Fromm, Michael E.
;  APPLICANT:  Heard, Jacqueline E.
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Adam, Luc J.
;  APPLICANT:  Broun, Pierre E.
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Reuber, T. Lynne
```

```
;  APPLICANT:  Keddie, James S.
;  APPLICANT:  Yu, Guo-Liang
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Samaha, Raymond R.
;  APPLICANT:  Pilgrim, Marsha L.
;  APPLICANT:  Creelman, Robert A.
;  APPLICANT:  DuBell, Arnold N.
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Kumimoto, Roderick
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
;  FILE REFERENCE: MBI-0048CIP
;  CURRENT APPLICATION NUMBER: US/10/412,699B
;  CURRENT FILING DATE:  2003-04-10
;  PRIOR APPLICATION NUMBER: 09/394,519
;  PRIOR FILING DATE: 1999-09-13
;  PRIOR APPLICATION NUMBER: 09/489,376
;  PRIOR FILING DATE: 2000-01-21
;  PRIOR APPLICATION NUMBER: 09/506,720
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 09/533,030
;  PRIOR FILING DATE: 2000-03-22
;  PRIOR APPLICATION NUMBER: 09/533,392
;  PRIOR FILING DATE: 2000-03-22
;  PRIOR APPLICATION NUMBER: 09/533,029
;  PRIOR FILING DATE: 2000-03-22
;  PRIOR APPLICATION NUMBER: 09/532,591
;  PRIOR FILING DATE: 2000-03-22
;  PRIOR APPLICATION NUMBER: 09/533,648
;  PRIOR FILING DATE: 2000-03-22
;  PRIOR APPLICATION NUMBER: 09/713,994
;  PRIOR FILING DATE: 2000-11-16
;  PRIOR APPLICATION NUMBER: 09/819,142
;  PRIOR FILING DATE: 2001-03-27
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2011
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1681
;   LENGTH: 289
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-10-412-699B-1681

  Query Match             81.0%;  Score 34;  DB 12;  Length 289;
  Best Local Similarity   57.1%;  Pred. No. 8.1e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db        279 CSCSPYP 285


RESULT 7
US-10-374-780A-1670
; Sequence 1670, Application US/10374780A
; Publication No. US20040019927A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, Bradley K
;  APPLICANT:  Riechmann, Jose Luis
;  APPLICANT:  Jiang, Cai-Zhong
;  APPLICANT:  Heard, Jacqueline E
;  APPLICANT:  Haake, Volker
;  APPLICANT:  Creelman, Robert A
;  APPLICANT:  Ratcliffe, Oliver
;  APPLICANT:  Adam, Luc J
;  APPLICANT:  Reuber, T. Lynne
;  APPLICANT:  Keddie, James
;  APPLICANT:  Broun, Pierre E
;  APPLICANT:  Pilgrim, Marsha L
;  APPLICANT:  Dubell III, Arnold T
;  APPLICANT:  Pineda, Omaira
;  APPLICANT:  Yu, Guo-Liang
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
;  FILE REFERENCE: MBI-0047 CIP
;  CURRENT APPLICATION NUMBER: US/10/374,780A
;  CURRENT FILING DATE:  2003-02-25
;  PRIOR APPLICATION NUMBER: 09/837,944
;  PRIOR FILING DATE: 2001-04-18
;  PRIOR APPLICATION NUMBER: 60/310,847
;  PRIOR FILING DATE: 2001-08-09
;  PRIOR APPLICATION NUMBER: 09/934,455
;  PRIOR FILING DATE: 2001-08-22
;  PRIOR APPLICATION NUMBER: 60/336,049
;  PRIOR FILING DATE: 2001-11-19
;  PRIOR APPLICATION NUMBER: 60/338,692
;  PRIOR FILING DATE: 2001-12-11
;  PRIOR APPLICATION NUMBER: 10/171,468
;  PRIOR FILING DATE: 2002-06-14
;  PRIOR APPLICATION NUMBER: 10/225,066
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: 10/225,067
;  PRIOR FILING DATE: 2002-08-09
;  PRIOR APPLICATION NUMBER: 10/225,068
;  PRIOR FILING DATE: 2002-08-09
;  NUMBER OF SEQ ID NOS: 2906
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1670
;   LENGTH: 289
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Orthologous to G1540
US-10-374-780A-1670

  Query Match             81.0%;  Score 34;  DB 15;  Length 289;
  Best Local Similarity   57.1%;  Pred. No. 8.1e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db        279 CSCSPYP 285


RESULT 8
US-10-424-599-153612
; Sequence 153612, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 153612
;   LENGTH: 91
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_109735C.1.pep
US-10-424-599-153612

  Query Match             78.6%;  Score 33;  DB 12;  Length 91;
  Best Local Similarity   66.7%;  Pred. No. 4.6e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         16 CSCSPH 21


RESULT 9
```

```
US-09-925-297-595
; Sequence 595, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA105
;  CURRENT APPLICATION NUMBER: US/09/925,297
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05989
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 928
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 595
;   LENGTH: 97
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (38)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-297-595

  Query Match             78.6%;  Score 33;  DB 9;  Length 97;
  Best Local Similarity   66.7%;  Pred. No. 4.9e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         46 CACSPH 51


RESULT 10
US-09-796-692-1071
; Sequence 1071, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1071
;   LENGTH: 108
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-796-692-1071

  Query Match             78.6%;  Score 33;  DB 9;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 5.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db          6 CSCSPH 11


RESULT 11
US-10-040-862-1071
; Sequence 1071, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-013520US
;  CURRENT APPLICATION NUMBER: US/10/040,862
;  CURRENT FILING DATE:  2001-11-06
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: US 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  PRIOR APPLICATION NUMBER: US 09/796,692
;  PRIOR FILING DATE: 2001-03-01
;  NUMBER OF SEQ ID NOS: 10467
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1071
;   LENGTH: 108
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-040-862-1071

  Query Match             78.6%;  Score 33;  DB 14;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 5.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db          6 CSCSPH 11
```

```
RESULT 12
US-10-057-475B-1071
; Sequence 1071, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Clapper, Jonathan David
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Ordonez, Nadia
;  APPLICANT:  Carter, Lauren
;  APPLICANT:  McNeill, Patricia Dianne
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-014402US
;  CURRENT APPLICATION NUMBER: US/10/057,475B
;  CURRENT FILING DATE:  2002-01-22
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 10979
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1071
;   LENGTH: 108
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-057-475B-1071

  Query Match             78.6%;  Score 33;  DB 15;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 5.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db          6 CSCSPH 11


RESULT 13
US-10-154-884B-1071
; Sequence 1071, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-013521US
;  CURRENT APPLICATION NUMBER: US/10/154,884B
;  CURRENT FILING DATE:  2002-05-23
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 11290
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1071
;   LENGTH: 108
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-154-884B-1071

  Query Match             78.6%;  Score 33;  DB 15;  Length 108;
  Best Local Similarity   66.7%;  Pred. No. 5.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db          6 CSCSPH 11


RESULT 14
US-09-796-692-1736
; Sequence 1736, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
;  TITLE OF INVENTION:  HEMATOLOGICAL MALIGNANCIES
;  FILE REFERENCE: 2077.001200
;  CURRENT APPLICATION NUMBER: US/09/796,692
;  CURRENT FILING DATE:  2001-03-01
;  PRIOR APPLICATION NUMBER: 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: 60/223,378
```

```
;  PRIOR FILING DATE: 2000-08-07
;  NUMBER OF SEQ ID NOS: 9597
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1736
;   LENGTH: 119
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variant
;   LOCATION: (1)...(119)
;   OTHER INFORMATION: Xaa = Any amino acid
US-09-796-692-1736

  Query Match             78.6%;  Score 33;  DB 9;  Length 119;
  Best Local Similarity   66.7%;  Pred. No. 5.7e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db          6 CSCSPH 11


RESULT 15
US-10-040-862-1736
; Sequence 1736, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
;  APPLICANT: Gaiger, Alexander
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Corixa Corporation
;  TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;  TITLE OF INVENTION:  Hematological Malignancies
;  FILE REFERENCE: 014058-013520US
;  CURRENT APPLICATION NUMBER: US/10/040,862
;  CURRENT FILING DATE:  2001-11-06
;  PRIOR APPLICATION NUMBER: US 60/186,126
;  PRIOR FILING DATE: 2000-03-01
;  PRIOR APPLICATION NUMBER: US 60/190,479
;  PRIOR FILING DATE: 2000-03-17
;  PRIOR APPLICATION NUMBER: US 60/200,545
;  PRIOR FILING DATE: 2000-04-27
;  PRIOR APPLICATION NUMBER: US 60/200,303
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,779
;  PRIOR FILING DATE: 2000-04-28
;  PRIOR APPLICATION NUMBER: US 60/200,999
;  PRIOR FILING DATE: 2000-05-01
;  PRIOR APPLICATION NUMBER: US 60/202,084
;  PRIOR FILING DATE: 2000-05-04
;  PRIOR APPLICATION NUMBER: US 60/206,201
;  PRIOR FILING DATE: 2000-05-22
;  PRIOR APPLICATION NUMBER: US 60/218,950
;  PRIOR FILING DATE: 2000-07-14
;  PRIOR APPLICATION NUMBER: US 60/222,903
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: US 60/223,416
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/223,378
;  PRIOR FILING DATE: 2000-08-07
;  PRIOR APPLICATION NUMBER: US 09/796,692
;  PRIOR FILING DATE: 2001-03-01
;  NUMBER OF SEQ ID NOS: 10467
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1736
;   LENGTH: 119
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variant
;   LOCATION: (1)...(119)
;   OTHER INFORMATION: Xaa = Any amino acid
US-10-040-862-1736

  Query Match             78.6%;  Score 33;  DB 14;  Length 119;
  Best Local Similarity   66.7%;  Pred. No. 5.7e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db          6 CSCSPH 11


Search completed: April  8, 2004, 11:55:41
Job time : 30.3333 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 8, 2004, 11:00:44 ; Search time 44.0417 Seconds
(without alignments)
44.908 Million cell updates/sec

Title: US-09-753-139C-1
Perfect score: 42
Sequence: 1 CXCXPHP 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	95.2	55	4	AAU14695	Aau14695 Novel bon
2	40	95.2	70	5	ABP06775	Abp06775 Human ORF
3	39	92.9	30	5	ABB79208	Abb79208 Human pro
4	38	90.5	7	5	ABB83462	Abb83462 Tissue In
5	38	90.5	73	4	AAU40466	Aau40466 Propionib
6	38	90.5	73	6	ABM36985	Abm36985 Propionib
7	38	90.5	93	5	ABP58502	Abp58502 Human sho
8	38	90.5	119	4	AAM06377	Aam06377 Human foe
9	38	90.5	221	4	AAB93924	Aab93924 Human pro
10	38	90.5	225	3	AAY95940	Aay95940 Porcine a
11	38	90.5	527	6	ABP72624	Abp72624 Prawn chi
12	37	88.1	122	4	AAU29832	Aau29832 Novel hum
13	37	88.1	170	5	ABB89097	Abb89097 Human pol
14	34	81.0	104	4	AAU49509	Aau49509 Propionib
15	34	81.0	104	6	ABM46028	Abm46028 Propionib
16	33	78.6	36	3	AAB44849	Aab44849 Human sec
17	33	78.6	86	4	ABG23045	Abg23045 Novel hum
18	33	78.6	97	3	AAB54143	Aab54143 Human pan
19	33	78.6	108	4	AAM80707	Aam80707 Human hae
20	33	78.6	119	4	AAM81372	Aam81372 Human hae
21	33	78.6	120	4	AAM80701	Aam80701 Human hae
22	33	78.6	139	4	AAM80391	Aam80391 Human hae
23	33	78.6	149	4	AAM80713	Aam80713 Human hae
24	33	78.6	155	4	AAM81278	Aam81278 Human hae
25	33	78.6	592	7	ADE86252	Ade86252 Hedgehog
26	32	76.2	40	2	AAY00299	Aay00299 Human sec
27	32	76.2	40	7	ADD90328	Add90328 Novel hum
28	32	76.2	64	4	AAU51994	Aau51994 Propionib
29	32	76.2	64	6	ABM48513	Abm48513 Propionib
30	32	76.2	80	5	ABP02671	Abp02671 Human ORF
31	32	76.2	94	4	ABG11465	Abg11465 Novel hum
32	32	76.2	110	4	ABB17019	Abb17019 Human ner
33	32	76.2	117	6	ABP76241	Abp76241 Human GEN
34	32	76.2	118	4	ABG24883	Abg24883 Novel hum
35	32	76.2	125	4	ABG14479	Abg14479 Novel hum
36	32	76.2	129	3	AAG03981	Aag03981 Human sec
37	32	76.2	134	6	ABU00091	Abu00091 Human nov
38	32	76.2	135	5	ABR01728	Abr01728 Human bre
39	32	76.2	148	4	ABG19243	Abg19243 Novel hum
40	32	76.2	149	3	AAB54086	Aab54086 Human pan
41	32	76.2	164	4	AAG75009	Aag75009 Human col
42	32	76.2	171	4	ABG20830	Abg20830 Novel hum
43	32	76.2	194	7	ADB64752	Adb64752 Human pro
44	32	76.2	215	4	AAB63957	Aab63957 Human pro
45	32	76.2	215	4	ABG13901	Abg13901 Novel hum

ALIGNMENTS

RESULT 1
AAU14695
ID AAU14695 standard; protein; 55 AA.
XX
AC AAU14695;
XX
DT 24-OCT-2001 (first entry)
XX
DE Novel bone marrow polypeptide #94.
XX
KW Bone marrow; diagnostic; therapeutic; gene therapy; antigenic;
KW haematopoiesis; myeloid; lymph cell disorder; tissue regeneration;
KW wound healing; nutritional supplement; immune disorder;
KW severe combined immunodeficiency; SCID.
XX
OS Homo sapiens.
XX
PN WO200157187-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US003782.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 30-NOV-2000; 2000US-0250683P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Zhou P, Xue AJ;
PI Ren F, Drmanac RT;
XX
DR WPI; 2001-488875/53.
DR N-PSDB; AAS23000.
XX
PT Nucleic acids encoding bone marrow polypeptides, useful in diagnostic and
PT gene therapy.
XX
PS Claim 10; Page 251-252; 392pp; English.
XX
CC AAU14602-AAU14794 represent novel bone marrow polypeptides of the
CC invention. The proteins and corresponding coding sequences may be used in
CC the prevention, diagnosis and treatment of diseases associated with
CC inappropriate bone marrow polypeptide expression. For example, to treat
CC disorders associated with decreased expression by rectifying mutations or
CC deletions in a patient's genome that affect the activity of the
CC polypeptides by expressing inactive proteins or to supplement the

CC patient's own production of the polypeptide. Additionally, the nucleic
CC acids may be used to produce the polypeptides, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein. The
CC nucleic acid and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acid sequences in samples, and therefore which patients
CC may be in need of restorative therapy. The proteins may also be used as
CC antigens in the production of antibodies against bone marrow proteins and
CC in assays to identify modulators of their expression and activity. The
CC anti-bone marrow protein antibodies and antagonists may also be used to
CC down regulate expression and activity. The antibodies may also be used as
CC diagnostic agents for detecting the presence of the protein in samples
CC (e.g. by enzyme linked immunosorbant assay (ELISA)). The proteins may be
CC used to regulate haematopoiesis activity, and consequently in the
CC treatment of myeloid or lymph cell disorders; in tissue regeneration,
CC such as wound healing; as a nutritional supplement; and in treatment of
CC immune disorders such as severe combined immunodeficiency (SCID)
XX
SQ Sequence 55 AA;

Query Match 95.2%; Score 40; DB 4; Length 55;
Best Local Similarity 71.4%; Pred. No. 41;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
| | |||
Db 10 CSCTPHP 16

RESULT 2
ABP06775
ID ABP06775 standard; protein; 70 AA.
XX
AC ABP06775;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:13532.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR N-PSDB; ABN22527.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 13532; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 70 AA;

Query Match 95.2%; Score 40; DB 5; Length 70;
Best Local Similarity 71.4%; Pred. No. 50;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
| | |||
Db 49 CSCTPHP 55

RESULT 3
ABB79208
ID ABB79208 standard; protein; 30 AA.
XX
AC ABB79208;
XX
DT 08-AUG-2002 (first entry)
XX
DE Human prostate specific protein sequence SEQ ID NO:157.
XX
KW Human; prostate specific gene; prostate specific protein; PSG; PSP;
KW prostate cancer.
XX
OS Homo sapiens.
XX
PN WO200236808-A2.
XX
PD 10-MAY-2002.
XX
PF 05-NOV-2001; 2001WO-US047283.
XX
PR 03-NOV-2000; 2000US-0245740P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Recipon H, Chen S, Liu C;
XX
DR WPI; 2002-471506/50.
XX
PT New prostate-specific nucleic acids and polypeptides, useful for
PT identifying, diagnosing, monitoring, staging, imaging, and treating
PT prostate cancer and non-cancerous disease states in prostate tissue.
XX
PS Claim 11; Page 223; 254pp; English.
XX
CC ABN87650 to ABN87789 represent human prostate-specific nucleic acids (I),
CC and ABB79192 to ABB79295 represent human prostate-specific proteins (II)
CC from the present invention. (I) and (II) have cytostatic activity. (I)

CC can be used in gene therapy. The prostate-specific nucleic acids,
CC polypeptides and compositions from the present invention can be used for
CC identifying, diagnosing, monitoring, staging, imaging, and treating
CC prostate cancer and non-cancerous disease states in prostate tissue; for
CC identifying prostate tissue; for monitoring, identifying and/or designing
CC agonists and antagonists of the polypeptides; in gene therapy; in
CC producing transgenic animals and cells; for producing engineered prostate
CC tissue for treatment and research; and as elements in an array or
CC computer program for pattern recognition of prostate disorders. The
CC nucleic acids may be used as hybridisation probes to detect, characterise
CC and quantify hybridising nucleic acids in, and isolate hybridising
CC nucleic acids from, both genomic and transcript-derived nucleic acid
CC samples
XX
SQ Sequence 30 AA;

Query Match 92.9%; Score 39; DB 5; Length 30;
Best Local Similarity 71.4%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
| | |||
Db 24 CACFPHP 30

RESULT 4
ABB83462
ID ABB83462 standard; peptide; 7 AA.
XX
AC ABB83462;
XX
DT 30-SEP-2002 (first entry)
XX
DE Tissue Inhibitor of Metalloproteinase, TIMP, derived peptide #1.
XX
KW MMP; Matrix Metalloproteinase; zinc chelator; chronic wound; acute wound;
KW Tissue Inhibitor of Metalloproteinase; TIMP; connective tissue breakdown;
KW angiogenesis-associated disorder.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 2
FT /label= Ser, Thr
FT Misc-difference 4
FT /label= Ser, Ala, Val
XX
PN WO200253173-A2.
XX
PD 11-JUL-2002.
XX
PF 21-DEC-2001; 2001WO-US049276.
XX
PR 29-DEC-2000; 2000US-00753139.
XX
PA (KIMB) KIMBERLY-CLARK WORLDWIDE INC.
XX
PI Quirk S, Tyrrell DJ;
XX
DR WPI; 2002-583595/62.
XX
PT New matrix metalloproteinase regulator useful for the treatment of
PT chronic and acute wounds comprises a zinc chelator and a tissue inhibitor
PT of metalloproteinases-derived peptide.
XX
PS Claim 5; Page 15; 57pp; English.
XX
CC The present invention relates to Matrix Metalloproteinase (MMP)
CC regulators, which comprise a zinc chelator and a Tissue Inhibitor of
CC Metalloproteinases (TIMP)-derived peptide. The present sequence is one
CC such TIMP-derived peptide used to generate the MMP regulators. The MMP
CC regulators are useful for treating chronic and acute wounds, angiogenesis
CC -associated disorders; and other diseases and disorders involving
CC uncontrolled breakdown of connective tissues by MMPs. MMPs contain a zinc
CC molecule located in the active site, which participates in degrading
CC collagen. The binding specificity of the TIMP-derived peptide brings the
CC zinc chelator into molecular proximity of the MMP bound zinc in such a
CC way to allow ligation. This results in the regulation of the level of MMP
CC activity to promote wound healing by providing a MMP regulator having
CC high affinity and selectivity
XX
SQ Sequence 7 AA;

Query Match 90.5%; Score 38; DB 5; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.4e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
|||||||
Db 1 CXCXPHP 7

RESULT 5
AAU40466
ID AAU40466 standard; protein; 73 AA.
XX
AC AAU40466;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #1362.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
DR N-PSDB; AAS59511.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 1661; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

```
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 73 AA;

  Query Match             90.5%;  Score 38;  DB 4;  Length 73;
  Best Local Similarity   71.4%;  Pred. No. 1e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         56 CRCIPHP 62


RESULT 6
ABM36985
ID   ABM36985 standard; protein; 73 AA.
XX
AC   ABM36985;
XX
DT   20-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes predicted ORF-encoded polypeptide #1661.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
DR   N-PSDB; ACF64440.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Example 1; SEQ ID NO 1661; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a polypeptide predicted to be encoded by an ORF (open
CC   reading frame) contained within the P. acnes polynucleotides of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 73 AA;

  Query Match             90.5%;  Score 38;  DB 6;  Length 73;
  Best Local Similarity   71.4%;  Pred. No. 1e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         56 CRCIPHP 62


RESULT 7
ABP58502
ID   ABP58502 standard; protein; 93 AA.
XX
AC   ABP58502;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Human short chain dehydrogenase 10.23.
XX
KW   Human; short chain dehydrogenase 10.23; recombinant production;
KW   gene therapy; malignant tumour; cancer; blood disease;
KW   human immunodeficiency virus; HIV infection; immune disorder;
KW   inflammatory condition; cytostatic; antiinflammatory; immunomodulator;
KW   enzyme.
XX
OS   Homo sapiens.
XX
PN   CN1358846-A.
XX
PD   17-JUL-2002.
XX
PF   13-DEC-2000; 2000CN-00127866.
XX
PR   13-DEC-2000; 2000CN-00127866.
XX
PA   (SHAN-) SHANGHAI BIOWINDOW GENE DEV INC.
XX
PI   Mao Y,  Xie Y;
XX
DR   WPI; 2002-733567/80.
DR   N-PSDB; ABV76533.
XX
PT   Novel polypeptide-human short chain dehydrogenase 10.23 the
PT   polynucleotide for encoding the polypeptide.
XX
PS   Claim 1; Page 26 (Disclosure); 32pp; Chinese.
XX
CC   The invention relates to human short chain dehydrogenase 10.23 (ABP58502)
CC   and nucleic acids encoding it (ABV76533). The protein has a molecular
CC   weight of 10.23 kD. The invention also relates to a method for the
CC   recombinant production of the protein, an antagonist of the protein, and
CC   the use of the protein, gene and antagonist in therapeutic applications.
CC   Short chain dehydrogenase 10.23 can be used in the treatment of a variety
CC   of diseases such as malignant tumours, blood diseases, HIV (human
CC   immunodeficiency virus) infection, immune disorders and inflammatory
CC   conditions. The present sequence represents human short chain
CC   dehydrogenase 10.23
XX
SQ   Sequence 93 AA;
```

```
  Query Match           90.5%;  Score 38;  DB 5;  Length 93;
  Best Local Similarity   71.4%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         22 CICRPHP 28


RESULT 8
AAM06377
ID   AAM06377 standard; protein; 119 AA.
XX
AC   AAM06377;
XX
DT   05-OCT-2001  (first entry)
XX
DE   Human foetal protein, SEQ ID NO: 108.
XX
KW   Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
KW   nootropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
KW   gene therapy; antisense therapy; cancer; immune disorder;
KW   growth disorder; osteoporosis; thrombolytic disorder;
KW   nervous system disorder; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200155339-A2.
XX
PD   02-AUG-2001.
XX
PF   25-JAN-2001; 2001WO-US002723.
XX
PR   25-JAN-2000; 2000US-00491404.
PR   15-SEP-2000; 2000US-00663870.
PR   06-NOV-2000; 2000US-00707351.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Yeung G,  Ford JE,  Boyle BJ,  Arterburn MC,  Drmanac RA,  Tang YT;
PI   Liu C,  Asundi V,  Zhou P,  Werhman T;
XX
DR   WPI; 2001-465571/50.
DR   N-PSDB; AAH94052.
XX
PT   Novel fetal proteins useful for the treatment and diagnosis of diseases
PT   associated with dysfunction of the protein e.g. cancers, immune
PT   disorders, growth disorders, thrombolytic disorders, nervous system
PT   disorders and inflammation.
XX
PS   Claim 10; Page 220; 715pp; English.
XX
CC   The invention relates to novel foetal polypeptides encoded by
CC   polynucleotides comprising one of 477 sequences fully defined in the
CC   specification. The foetal polynucleotides and polypeptides are useful in
CC   the treatment and diagnosis of diseases such as cancers, immune
CC   disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,
CC   nervous system disorders and inflammation. The present sequence is a
CC   polypeptide encoded by a cDNA assembled using an expressed sequence tag
CC   (EST) found to be expressed in human foetal tissue cDNA libraries
XX
SQ   Sequence 119 AA;

  Query Match           90.5%;  Score 38;  DB 4;  Length 119;
  Best Local Similarity   71.4%;  Pred. No. 1.5e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         13 CRCVPHP 19


RESULT 9
AAB93924
ID   AAB93924 standard; protein; 221 AA.
XX
AC   AAB93924;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:13913.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 13913; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 221 AA;

  Query Match           90.5%;  Score 38;  DB 4;  Length 221;
  Best Local Similarity   71.4%;  Pred. No. 2.4e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db         13 CRCVPHP 19
```

```
RESULT 10
AAY95940
ID   AAY95940 standard; protein; 225 AA.
XX
AC   AAY95940;
XX
DT   12-SEP-2003  (revised)
DT   20-NOV-2000  (first entry)
XX
DE   Porcine adenovirus 3 IVa2 protein.
XX
KW   PAd3; adenovirus; mastadenovirus; vector; IVa2 protein.
XX
OS   Porcine adenovirus 3.
XX
PN   WO200050076-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US004711.
XX
PR   24-FEB-1999;   99US-0121647P.
XX
PA   (PURD ) PURDUE RES FOUND.
XX
PI   Mittal SK,  Aggarwal N;
XX
DR   WPI; 2000-572039/53.
DR   N-PSDB; AAA50439.
XX
PT   Isolated nucleic acid molecules encoding proteins from porcine
PT   adenovirus, used for controlling viral replication, especially in the
PT   latent viral stage.
XX
PS   Example 1; Fig 1G; 77pp; English.
XX
CC   The present sequence is that of the protein encoded by an open reading
CC   frame, IVa2, identified in a porcine adenovirus 3 (PAd3) genome fragment
CC   (see AAA50439). IVa2 shows 77% and 72% amino acid identity with IVa2
CC   proteins of human adenovirus 5 and bovine adenovirus 3, respectively. The
CC   invention provides recombinant adenovirus vectors including novel open
CC   reading frame 97R, 162R, 163R*, 288R or 184R and at least 1 foreign
CC   promoter. In a preferred form of the invention, a plasmid is provided
CC   that includes the nucleotide sequence encoding the novel PAd3 protein,
CC   along with an adenoviral origin of replication. A replication-defective
CC   PAd3 vector comprises a PAd3 genome having a functional deletion in the
CC   E1 region, and with a nucleotide sequence encoding a protein of interest
CC   inserted into the E1 region. Methods for making an adenovirus vector and
CC   for expressing a desired protein or a viral protein in a host cell are
CC   claimed. (Updated on 12-SEP-2003 to standardise OS field)
XX
SQ   Sequence 225 AA;

  Query Match             90.5%;  Score 38;  DB 3;  Length 225;
  Best Local Similarity   71.4%;  Pred. No. 2.4e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db        169 CTCPPHP 175


RESULT 11
ABP72624
ID   ABP72624 standard; protein; 527 AA.
XX
AC   ABP72624;
XX
DT   23-OCT-2003  (revised)
DT   11-JUN-2003  (first entry)
XX
DE   Prawn chitinase.
XX
KW   Insecticide; pesticide; insect control; insect; toxin; chitinase; enzyme;
KW   neuropeptide; transgenic plant; crop protection; prawn.
XX
OS   Marsupenaeus japonicus.
XX
PN   WO2003014150-A2.
XX
PD   20-FEB-2003.
XX
PF   06-AUG-2002; 2002WO-GB003598.
XX
PR   08-AUG-2001; 2001GB-00019274.
XX
PA   (UYDU-) UNIV DURHAM.
PA   (ENVI-) DEPT ENVIRONMENT FOOD & RURAL AFFAIRS.
XX
PI   Gatehouse JA,  Fitches EC,  Edwards JP;
XX
DR   WPI; 2003-278469/27.
XX
PT   Fusion protein useful for combating insect pests, comprises a
PT   translocating moiety comprising a plant protein capable of acting as a
PT   carrier to translocate toxic moiety inside plant pathogen, and a toxic
PT   moiety.
XX
PS   Claim 7; Fig 7; 51pp; English.
XX
CC   The present sequence is that of prawn (Penaeus japonicus) chitinase. This
CC   protein can be used in claimed fusion proteins of the invention
CC   comprising a translocating moiety and a toxic moiety, where the
CC   translocating moiety is a plant protein (e.g. a lectin) capable of acting
CC   as a carrier to translocate the toxic moiety across the gut wall of a
CC   plant pathogen, and the toxic moiety is an arthropod-derived peptide or
CC   protein capable of causing deleterious effects on growth, development,
CC   reproduction or mortality in pest insects. Suitable arthropod peptides
CC   and proteins include allatostatins, chitinase, diuretic hormone and their
CC   metabolites and analogues. Polynucleotides encoding the fusion protein,
CC   vectors, host cells and transgenic plants that are resistant to disease
CC   are also provided. The fusion protein is target-specific, and resists
CC   degradation in the insect gut. (Updated on 23-OCT-2003 to standardise OS
CC   field)
XX
SQ   Sequence 527 AA;

  Query Match             90.5%;  Score 38;  DB 6;  Length 527;
  Best Local Similarity   71.4%;  Pred. No. 4.7e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |||
Db        500 CGCEPHP 506


RESULT 12
AAU29832
ID   AAU29832 standard; protein; 122 AA.
XX
AC   AAU29832;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Novel human secreted protein #323.
XX
KW   Human; vaccination; gene therapy; nutritional supplement;
KW   stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW   immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO200179449-A2.
XX
PD   25-OCT-2001.
```

XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 198; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 122 AA;

Query Match 88.1%; Score 37; DB 4; Length 122;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
| | |||
Db 80 CPCLPHP 86

RESULT 13
ABB89097
ID ABB89097 standard; protein; 170 AA.
XX
AC ABB89097;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 1473.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
PN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US016450.
XX
PR 19-MAY-2000; 2000US-0205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
DR N-PSDB; ABL89506.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
PS Claim 11; SEQ ID NO 1473; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 170 AA;

Query Match 88.1%; Score 37; DB 5; Length 170;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
| | |||
Db 114 CLCPPHP 120

RESULT 14
AAU49509
ID AAU49509 standard; protein; 104 AA.
XX
AC AAU49509;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #10405.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

```
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59545.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris.
XX
PS   Example 1; SEQ ID NO 10704; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention
CC   and determining the amount of bound protein in the sample. The
CC   polypeptides may be used as antigens in the production of antibodies
CC   specific for P. acnes proteins. These antibodies can be used to
CC   downregulate expression and activity of P. acnes polypeptides and
CC   therefore treat P. acnes infections. The antibodies may also be used as
CC   diagnostic agents for determining P. acnes presence, for example, by
CC   enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC   this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 104 AA;

  Query Match             81.0%;  Score 34;  DB 4;  Length 104;
  Best Local Similarity   57.1%;  Pred. No. 5.1e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db         96 CSCSPYP 102


RESULT 15
ABM46028
ID   ABM46028 standard; protein; 104 AA.
XX
AC   ABM46028;
XX
DT   20-OCT-2003  (first entry)
XX
DE   Propionibacterium acnes predicted ORF-encoded polypeptide #10704.
XX
KW   Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW   immunostimulant; immune response; vaccine.
XX
OS   Propionibacterium acnes.
XX
PN   WO2003033515-A1.
XX
PD   24-APR-2003.
XX
PF   11-OCT-2002; 2002WO-US032727.
XX
PR   15-OCT-2001; 2001US-00978825.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Mitcham JL,  Skeiky YAW,  Persing DH,  Bhatia A,  Maisonneuve JL;
PI   Zhang Y,  Wang S,  Jen S,  Lodes MJ,  Benson DR,  Jones R,  Carter D;
PI   Barth B,  Vallieve-Douglass J;
XX
DR   WPI; 2003-381789/36.
DR   N-PSDB; ACF64474.
XX
PT   New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT   polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT   or for stimulating an immune response specific for a P. acnes protein.
XX
PS   Example 1; SEQ ID NO 10704; 1481pp; English.
XX
CC   The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC   encoding a Propionibacterium acnes protein. The invention also relates to
CC   polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC   immunogenic fragments of P. acnes polypeptides. The invention
CC   additionally encompasses expression vectors and host cells comprising a
CC   polynucleotide of the invention; antibodies against polypeptides of the
CC   invention; fusion proteins comprising a polypeptide of the invention; a
CC   method for stimulating an immune response specific for a P. acnes
CC   polypeptide and an isolated T cell population comprising T cells prepared
CC   via this method; a vaccine composition (comprising P. acnes polypeptides,
CC   polynucleotides, antibodies, fusion proteins, T cell populations, or
CC   antigen-presenting cells that express the polypeptide); a method and kit
CC   for detecting or determining the presence or absence of P. acnes in a
CC   patient; and a method for inhibiting the development of P. acnes in a
CC   patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC   proteins, T cell populations or antigen-presenting cells that express the
CC   polypeptides are useful for diagnosing, preventing or treating acne
CC   vulgaris, or for stimulating an immune response specific for a P. acnes
CC   protein. The polynucleotides can also be used as probes or primers for
CC   nucleic acid hybridisation. The vaccine composition is useful for the
CC   stimulation of an immune response against P. acnes, or for treating acne,
CC   and the kit is useful for performing a diagnostic assay. The present
CC   sequence represents a polypeptide predicted to be encoded by an ORF (open
CC   reading frame) contained within the P. acnes polynucleotides of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 104 AA;

  Query Match             81.0%;  Score 34;  DB 6;  Length 104;
  Best Local Similarity   57.1%;  Pred. No. 5.1e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db         96 CSCSPYP 102


Search completed: April  8, 2004, 11:09:54
Job time : 46.1417 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 8, 2004, 11:02:14 ; Search time 12.7361 Seconds
(without alignments)
28.375 Million cell updates/sec

Title: US-09-753-139C-1
Perfect score: 42
Sequence: 1 CXCXPHP 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34	81.0	449	4	US-09-252-991A-17669	Sequence 17669, A
2	33	78.6	181	4	US-09-252-991A-30203	Sequence 30203, A
3	33	78.6	592	4	US-08-933-711B-18	Sequence 18, Appl
4	32	76.2	133	4	US-09-489-039A-8115	Sequence 8115, Ap
5	32	76.2	271	1	US-08-152-019A-28	Sequence 28, Appl
6	32	76.2	278	2	US-08-460-309-13	Sequence 13, Appl
7	32	76.2	278	2	US-08-125-077-13	Sequence 13, Appl
8	32	76.2	279	1	US-08-152-019A-29	Sequence 29, Appl
9	32	76.2	501	4	US-09-252-991A-18409	Sequence 18409, A
10	32	76.2	549	3	US-09-245-041-9	Sequence 9, Appli
11	32	76.2	1147	1	US-08-144-121-3	Sequence 3, Appli
12	32	76.2	1147	2	US-08-735-893-3	Sequence 3, Appli
13	32	76.2	1165	1	US-08-144-121-2	Sequence 2, Appli
14	32	76.2	1165	2	US-08-735-893-2	Sequence 2, Appli
15	32	76.2	1170	4	US-09-561-709B-12	Sequence 12, Appl
16	32	76.2	1172	4	US-09-919-172-16	Sequence 16, Appl
17	32	76.2	1260	3	US-09-245-041-2	Sequence 2, Appli
18	32	76.2	1587	4	US-09-845-583A-10	Sequence 10, Appl
19	32	76.2	1587	4	US-09-561-709B-3	Sequence 3, Appli
20	32	76.2	1725	4	US-09-562-702A-20	Sequence 20, Appl
21	32	76.2	1725	4	US-09-561-818A-20	Sequence 20, Appl
22	32	76.2	1765	4	US-09-562-702A-16	Sequence 16, Appl
23	32	76.2	1765	4	US-09-561-818A-16	Sequence 16, Appl
24	32	76.2	1786	4	US-09-562-702A-14	Sequence 14, Appl
25	32	76.2	1786	4	US-09-562-702A-18	Sequence 18, Appl
26	32	76.2	1786	4	US-09-561-818A-14	Sequence 14, Appl
27	32	76.2	1786	4	US-09-561-818A-18	Sequence 18, Appl
28	32	76.2	1786	4	US-09-561-709B-9	Sequence 9, Appli
29	31	73.8	7	1	US-08-179-481-42	Sequence 42, Appl
30	31	73.8	39	1	US-08-096-942-1	Sequence 1, Appli
31	31	73.8	39	5	PCT-US94-08063-1	Sequence 1, Appli
32	31	73.8	52	3	US-09-346-860-3	Sequence 3, Appli
33	31	73.8	52	4	US-09-735-685-3	Sequence 3, Appli
34	31	73.8	65	4	US-09-461-325-457	Sequence 457, App
35	31	73.8	65	4	US-10-012-542-457	Sequence 457, App
36	31	73.8	70	4	US-09-540-236-2438	Sequence 2438, Ap
37	31	73.8	116	3	US-09-081-320-29	Sequence 29, Appl
38	31	73.8	116	4	US-09-574-141A-29	Sequence 29, Appl
39	31	73.8	116	4	US-09-707-780-29	Sequence 29, Appl
40	31	73.8	117	3	US-09-081-320-7	Sequence 7, Appli
41	31	73.8	117	3	US-09-081-320-18	Sequence 18, Appl
42	31	73.8	117	4	US-09-574-141A-7	Sequence 7, Appli
43	31	73.8	117	4	US-09-574-141A-18	Sequence 18, Appl
44	31	73.8	117	4	US-09-574-141A-74	Sequence 74, Appl
45	31	73.8	117	4	US-09-707-780-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-252-991A-17669
; Sequence 17669, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17669
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17669

Query Match 81.0%; Score 34; DB 4; Length 449;
Best Local Similarity 57.1%; Pred. No. 3.2e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPHP 7
              | | |:|
Db         32 CSCSPYP 38
```

RESULT 2
US-09-252-991A-30203
; Sequence 30203, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30203
; LENGTH: 181
; TYPE: PRT

```
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30203

  Query Match             78.6%;  Score 33;  DB 4;  Length 181;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         91 CACAPH 96


RESULT 3
US-08-933-711B-18
; Sequence 18, Application US/08933711B
; Patent No. 6514724
; GENERAL INFORMATION:
;  APPLICANT: McMahon, Andrew P.
;  APPLICANT:  Chuang, Pao-Tien
;  TITLE OF INVENTION: HEDGEHOG INTERACTING PROTEINS AND USES RELATED THERETO
;  FILE REFERENCE: HUV-024.01
;  CURRENT APPLICATION NUMBER: US/08/933,711B
;  CURRENT FILING DATE:  1997-09-19
;  PRIOR APPLICATION NUMBER: 60/026,155
;  PRIOR FILING DATE: 1996-09-20
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Consensus
;   OTHER INFORMATION: sequence
US-08-933-711B-18

  Query Match             78.6%;  Score 33;  DB 4;  Length 592;
  Best Local Similarity   66.7%;  Pred. No. 5.7e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         73 CACSPH 78


RESULT 4
US-09-489-039A-8115
; Sequence 8115, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 8115
;   LENGTH: 133
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8115

  Query Match             76.2%;  Score 32;  DB 4;  Length 133;
  Best Local Similarity   66.7%;  Pred. No. 2.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         97 CGCSPH 102


RESULT 5
US-08-152-019A-28
; Sequence 28, Application US/08152019A
; Patent No. 5565331
;  GENERAL INFORMATION:
;    APPLICANT:  Tessier-Lavigne, Marc
;    APPLICANT:  Serafini, Tito
;    APPLICANT:  Kennedy, Timothy
;    APPLICANT:  Placzek, Marysia
;    APPLICANT:  Jessell, Thomas
;    APPLICANT:  Dodd, Jane
;    TITLE OF INVENTION:  NEURAL AXON OUTGROWTH MODULATORS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/152,019A
;      FILING DATE:  12-NOV-1993
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59012/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299 FHT UR
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  271 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-152-019A-28

  Query Match             76.2%;  Score 32;  DB 1;  Length 271;
  Best Local Similarity   66.7%;  Pred. No. 4.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        260 CSCLPH 265


RESULT 6
US-08-460-309-13
; Sequence 13, Application US/08460309
; Patent No. 5837496
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/460,309
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/125,077
;      FILING DATE:  22-SEP-1993
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  278 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-460-309-13

  Query Match             76.2%;  Score 32;  DB 2;  Length 278;
  Best Local Similarity   66.7%;  Pred. No. 4.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        259 CSCRPH 264


RESULT 7
US-08-125-077-13
; Sequence 13, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
;  GENERAL INFORMATION:
;    APPLICANT:  Engvall, Eva
;    APPLICANT:  Leivo, Ilmo
;    TITLE OF INVENTION:  Nucleic Acids Encoding Merosin, Merosin
;    TITLE OF INVENTION:  Fragments and Uses Thereof
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/125,077
;      FILING DATE:  22-SEP-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/US 94/10730
;      FILING DATE:  21-SEP-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/472,319
;      FILING DATE:  30-JAN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/919,951
;      FILING DATE:  27-JUL-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LA 9721
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  278 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-125-077-13

  Query Match             76.2%;  Score 32;  DB 2;  Length 278;
  Best Local Similarity   66.7%;  Pred. No. 4.3e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        259 CSCRPH 264


RESULT 8
US-08-152-019A-29
; Sequence 29, Application US/08152019A
; Patent No. 5565331
;  GENERAL INFORMATION:
;    APPLICANT:  Tessier-Lavigne, Marc
;    APPLICANT:  Serafini, Tito
;    APPLICANT:  Kennedy, Timothy
;    APPLICANT:  Placzek, Marysia
;    APPLICANT:  Jessell, Thomas
;    APPLICANT:  Dodd, Jane
;    TITLE OF INVENTION:  NEURAL AXON OUTGROWTH MODULATORS
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/152,019A
;      FILING DATE:  12-NOV-1993
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Osman, Richard Aron
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  A-59012/RAO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 398-3249
;      TELEX:  910 277299 FHT UR
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  279 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
```

US-08-152-019A-29

Query Match 76.2%; Score 32; DB 1; Length 279;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPH 6
| | ||
Db 260 CSCRPH 265

RESULT 9
US-09-252-991A-18409
; Sequence 18409, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18409
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18409

Query Match 76.2%; Score 32; DB 4; Length 501;
Best Local Similarity 57.1%; Pred. No. 7e+02;
Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CXCXPHP 7
| | | |
Db 132 CTCTPRP 138

RESULT 10
US-09-245-041-9
; Sequence 9, Application US/09245041
; Patent No. 6274339
; GENERAL INFORMATION:
; APPLICANT: Moore, K.
; APPLICANT: Nagle, D.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: OF BODY WEIGHT DISORDERS INCLUDING OBESITY
; FILE REFERENCE: 7853-136
; CURRENT APPLICATION NUMBER: US/09/245,041
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/093,630
; EARLIER FILING DATE: 1998-07-21
; EARLIER APPLICATION NUMBER: 60/104,978
; EARLIER FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: SITE
; LOCATION: all Xaa positions
; OTHER INFORMATION: Xaa=unknown amino acid
US-09-245-041-9

Query Match 76.2%; Score 32; DB 3; Length 549;
Best Local Similarity 66.7%; Pred. No. 7.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPH 6
| | ||
Db 138 CSCFPH 143

RESULT 11
US-08-144-121-3
; Sequence 3, Application US/08144121
; Patent No. 5610031
; GENERAL INFORMATION:
; APPLICANT: Burgeson, Robert E.
; APPLICANT: Wagman, David W.
; TITLE OF INVENTION: B1k CHAIN OF LAMININ AND METHODS OF USE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: BOSTON
; STATE: Massachusetts
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/144,121
; FILING DATE: 27-OCT-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Paul L.
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: (MGH-0780.0) MGP-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 1..231
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 232..411
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 412..765
; FEATURE:
; NAME/KEY: Domain
; LOCATION: 766..1147
US-08-144-121-3

Query Match 76.2%; Score 32; DB 1; Length 1147;
Best Local Similarity 66.7%; Pred. No. 1.4e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPH 6
| | ||
Db 461 CACDPH 466

RESULT 12
US-08-735-893-3
; Sequence 3, Application US/08735893
; Patent No. 5914317
; GENERAL INFORMATION:

```
;    APPLICANT:  Burgeson, Robert E.
;    APPLICANT:  Wagman, David W.
;    TITLE OF INVENTION:  B1k CHAIN OF LAMININ AND METHODS OF USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  BOSTON
;      STATE:  Massachusetts
;      COUNTRY:  United States
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/735,893
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/144,121
;      FILING DATE:  27-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  (MGH-0780.1) MGP-021DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1147 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    FRAGMENT TYPE:  internal
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  1..231
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  232..411
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  412..765
;    FEATURE:
;      NAME/KEY:  Domain
;      LOCATION:  766..1147
US-08-735-893-3

  Query Match             76.2%;  Score 32;  DB 2;  Length 1147;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        461 CACDPH 466


RESULT 13
US-08-144-121-2
; Sequence 2, Application US/08144121
; Patent No. 5610031
;  GENERAL INFORMATION:
;    APPLICANT:  Burgeson, Robert E.
;    APPLICANT:  Wagman, David W.
;    TITLE OF INVENTION:  B1k CHAIN OF LAMININ AND METHODS OF USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  BOSTON
;      STATE:  Massachusetts
;      COUNTRY:  United States
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/144,121
;      FILING DATE:  27-OCT-1993
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  (MGH-0780.0) MGP-021
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1165 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-144-121-2

  Query Match             76.2%;  Score 32;  DB 1;  Length 1165;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        478 CACDPH 483


RESULT 14
US-08-735-893-2
; Sequence 2, Application US/08735893
; Patent No. 5914317
;  GENERAL INFORMATION:
;    APPLICANT:  Burgeson, Robert E.
;    APPLICANT:  Wagman, David W.
;    TITLE OF INVENTION:  B1k CHAIN OF LAMININ AND METHODS OF USE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, suite 510
;      CITY:  BOSTON
;      STATE:  Massachusetts
;      COUNTRY:  United States
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/735,893
;      FILING DATE:  18-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/144,121
;      FILING DATE:  27-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Myers, Paul L.
;      REGISTRATION NUMBER:  35,965
;      REFERENCE/DOCKET NUMBER:  (MGH-0780.1) MGP-021DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
```

```
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1165 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-735-893-2

  Query Match             76.2%;  Score 32;  DB 2;  Length 1165;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        478 CACDPH 483


RESULT 15
US-09-561-709B-12
; Sequence 12, Application US/09561709B
; Patent No. 6682911
; GENERAL INFORMATION:
;  APPLICANT: Burgeson, Robert
;  APPLICANT:  Champliaud, Marie-France
;  APPLICANT:  Olson, Pamela
;  APPLICANT:  Koch, Manuel
;  APPLICANT:  Brunken, William
;  TITLE OF INVENTION: LAMININS AND USES THEREOF
;  FILE REFERENCE: 10287-060001
;  CURRENT APPLICATION NUMBER: US/09/561,709B
;  CURRENT FILING DATE:  2000-05-01
;  PRIOR APPLICATION NUMBER: US 09/168,949
;  PRIOR FILING DATE: 1998-10-09
;  PRIOR APPLICATION NUMBER: US 60/061,609
;  PRIOR FILING DATE: 1997-10-10
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 1170
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-561-709B-12

  Query Match             76.2%;  Score 32;  DB 4;  Length 1170;
  Best Local Similarity   66.7%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        479 CACDPH 484


Search completed: April  8, 2004, 11:20:27
Job time : 13.7361 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 8, 2004, 11:01:14 ; Search time 9.52778 Seconds
(without alignments)
70.671 Million cell updates/sec

Title: US-09-753-139C-1
Perfect score: 42
Sequence: 1 CXCXPHP 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	33	78.6	158	2	T04875	hypothetical prote
2	33	78.6	215	2	T19056	hypothetical prote
3	32	76.2	103	2	H72753	hypothetical prote
4	32	76.2	119	2	E89101	protein F25E5.9 [i
5	32	76.2	1170	2	A53612	laminin B1k chain
6	32	76.2	1376	1	VGIHJ2	E2 glycoprotein pr
7	32	76.2	1376	1	JQ1534	E2 glycoprotein pr
8	32	76.2	1786	1	MMHUB1	laminin beta-1 cha
9	32	76.2	1786	1	MMMSB1	laminin beta-1 cha
10	31	73.8	39	1	A48523	margatoxin [valida
11	31	73.8	68	2	JV0044	hypothetical 7.5K
12	31	73.8	87	2	A35666	transcription acti
13	31	73.8	119	1	CUPSAM	amicyanin precurso
14	31	73.8	131	1	S12972	amicyanin - Paraco
15	31	73.8	132	1	A23706	amicyanin - Thioba
16	31	73.8	145	2	T36527	hypothetical prote
17	31	73.8	149	2	F87579	hypothetical prote
18	31	73.8	211	2	T04119	probable serine/th
19	31	73.8	212	2	S53257	e antigen precurso
20	31	73.8	235	2	JC4603	conserved hypothet
21	31	73.8	253	2	S40181	phosphoprotein pho
22	31	73.8	253	2	T25768	hypothetical prote
23	31	73.8	282	2	T15640	hypothetical prote
24	31	73.8	313	1	FOVDA	gag polyprotein -
25	31	73.8	317	1	WMXR3S	34K protein - simi
26	31	73.8	317	1	WMXR3B	33K protein - bovi
27	31	73.8	317	1	WMXR5U	34K protein - porc
28	31	73.8	317	2	S30581	gp33 protein - hum
29	31	73.8	317	2	S49005	non-structural pro
30	31	73.8	343	2	T29547	hypothetical prote
31	31	73.8	352	2	T24369	hypothetical prote
32	31	73.8	355	2	T15797	hypothetical prote
33	31	73.8	448	2	T15188	hypothetical prote
34	31	73.8	454	2	T03130	probable tyrosine
35	31	73.8	458	2	S12444	hypothetical prote
36	31	73.8	459	2	F71257	hypothetical prote
37	31	73.8	486	1	JC7241	phosphoprotein pho
38	31	73.8	502	1	JC1283	phosphoprotein pho
39	31	73.8	506	2	A54190	cerebroside-sulfat
40	31	73.8	509	1	OKHULK	protein-tyrosine k
41	31	73.8	513	2	A38193	phosphoprotein pho
42	31	73.8	514	1	A36222	phosphoprotein pho
43	31	73.8	515	2	JT0976	phosphoprotein pho
44	31	73.8	518	2	A40942	phosphoprotein pho
45	31	73.8	521	1	S35067	phosphoprotein pho

ALIGNMENTS

RESULT 1
T04875
hypothetical protein F18F4.10 - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T04875
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15388
A;Accession: T04875
A;Molecule type: DNA
A;Residues: 1-158 <BEV>
A;Cross-references: EMBL:AL021637
A;Experimental source: cultivar Columbia; BAC clone F18F4
C;Genetics:
A;Map position: 4
A;Introns: 18/1; 72/1; 121/3
A;Note: F18F4.10

Query Match 78.6%; Score 33; DB 2; Length 158;
Best Local Similarity 57.1%; Pred. No. 1e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPHP 7
              | | |:|
Db         70 CTCTPNP 76
```

RESULT 2
T19056
hypothetical protein C07E3.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T19056
R;Matthews, P.
submitted to the EMBL Data Library, June 1995
A;Reference number: Z19066
A;Accession: T19056
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-215 <WIL>
A;Cross-references: EMBL:Z49908; PIDN:CAA90100.1; GSPDB:GN00020; CESP:C07E3.6
A;Experimental source: clone C07E3
C;Genetics:
A;Gene: CESP:C07E3.6
A;Map position: 2
A;Introns: 28/3; 136/3; 162/3; 190/3
C;Superfamily: Caenorhabditis elegans hypothetical protein C07E3.6

Query Match 78.6%; Score 33; DB 2; Length 215;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPH 6
              | | ||
Db         49 CSCTPH 54
```

RESULT 3
H72753
hypothetical protein APE0021 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C;Accession: H72753
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: H72753
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-103 <KAW>
A;Cross-references: DDBJ:AP000058; NID:g5103388; PIDN:BAA78930.1; PID:d1042706; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE0021

  Query Match 76.2%; Score 32; DB 2; Length 103;
  Best Local Similarity 57.1%; Pred. No. 1e+02;
  Matches 4; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 CXCXPHP 7
              | |  ||
Db         13 CSCTTHP 19
```

RESULT 4
E89101
protein F25E5.9 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C;Accession: E89101
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_ele
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A;Accession: E89101
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-119 <STO>
A;Cross-references: GB:chr_V; PIDN:AAC27339.1; PID:g3335246; GSPDB:GN00023; CESP:F25E5.9
C;Genetics:
A;Gene: F25E5.9
A;Map position: 5

  Query Match 76.2%; Score 32; DB 2; Length 119;
  Best Local Similarity 66.7%; Pred. No. 1.2e+02;
  Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPH 6
              | | ||
Db         71 CMCAPH 76
```

RESULT 5
A53612
laminin B1k chain precursor - human
N;Alternate names: kalinin B1 chain; nicein B1 chain
C;Species: Homo sapiens (man)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 05-Nov-1999
C;Accession: A53612
R;Gerecke, D.R.; Wagman, D.W.; Champliaud, M.F.; Burgeson, R.E.
J. Biol. Chem. 269, 11073-11080, 1994
A;Title: The complete primary structure for a novel laminin chain, the laminin B1k chain
A;Reference number: A53612; MUID:94209274; PMID:7512558
A;Accession: A53612
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1170 <GER>
A;Cross-references: GB:L25541; NID:g510702; PIDN:AAA61834.1; PID:g510703
A;Note: authors translated the codon CGA for residue 124 as Gln, GAG for residue 439 as .
C;Superfamily: laminin-type EGF-like homology
C;Keywords: glycoprotein
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-1170/Product: laminin B1k chain #status predicted <MAT>
F;250-312/Domain: laminin-type EGF-like homology <LEG1>
F;378-427/Domain: laminin-type EGF-like homology <LEG>
F;430-476/Domain: laminin-type EGF-like homology <EG11>
F;532-576/Domain: laminin-type EGF-like homology <LEG2>

  Query Match 76.2%; Score 32; DB 2; Length 1170;
  Best Local Similarity 66.7%; Pred. No. 6.8e+02;
  Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPH 6
              | | ||
Db        479 CACDPH 484
```

RESULT 6
VGIHJ2
E2 glycoprotein precursor - murine hepatitis virus (strain wild type MHV-4)
N;Alternate names: peplomer glycoprotein; spike glycoprotein
N;Contains: 90A glycoprotein; 90B glycoprotein
C;Species: murine hepatitis virus, MHV
C;Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 16-Jul-1999
C;Accession: A33748
R;Parker, S.E.; Gallagher, T.M.; Buchmeier, M.J.
Virology 173, 664-673, 1989
A;Title: Sequence analysis reveals extensive polymorphism and evidence of deletions with
A;Reference number: A33748; MUID:90085815; PMID:2556846
A;Accession: A33748
A;Molecule type: genomic RNA
A;Residues: 1-1376 <PAR>
A;Cross-references: GB:M32789; NID:g331846; PIDN:AAA46456.1; PID:g331847
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-14/Domain: signal sequence #status predicted <SIG>
F;15-1376/Product: E2 glycoprotein #status predicted <E2G>
F;15-769/Product: 90B glycoprotein #status predicted <EGB>
F;770-1376/Product: 90A glycoprotein #status predicted <EGA>
F;1321-1338/Domain: transmembrane #status predicted <TMN>
F;31,60,134,192,357,435,442,582,677,709,717,740,789,806,896,945,1178,1232,1242,1261,1277

  Query Match 76.2%; Score 32; DB 1; Length 1376;
  Best Local Similarity 57.1%; Pred. No. 7.7e+02;
  Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 CXCXPHP 7
              | | |:|
Db        532 CTCKPNP 538
```

RESULT 7
JQ1534
E2 glycoprotein precursor - murine hepatitis virus (strain JHM cl-2)
N;Alternate names: peplomer glycoprotein; spike glycoprotein
N;Contains: 90A glycoprotein; 90B glycoprotein
C;Species: murine hepatitis virus, MHV
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 07-May-1999
C;Accession: JQ1534
R;Taguchi, F.; Ikeda, T.; Shida, H.
J. Gen. Virol. 73, 1065-1072, 1992
A;Title: Molecular cloning and expression of a spike protein of neurovirulent murine cor
A;Reference number: JQ1534; MUID:92268864; PMID:1316938

```
A;Accession: JQ1534
A;Molecule type: mRNA
A;Residues: 1-1376 <TAG>
A;Cross-references: DDBJ:D10255
A;Note: the authors translated the codon TTT for residue 8 as Leu, GGG for residue 14 as
C;Superfamily: coronavirus E2 glycoprotein
C;Keywords: glycoprotein; transmembrane protein
F;1-14/Domain: signal sequence #status predicted <SIG>
F;15-1376/Product: E2 glycoprotein #status predicted <E2G>
F;15-769/Product: 90B glycoprotein #status predicted <EGB>
F;770-1376/Product: 90A glycoprotein #status predicted <EGA>
F;1318-1339/Domain: transmembrane #status predicted <TMM>
F;31,60,134,192,357,435,677,709,717,789,806,945,1232,1242,1261,1277,1298/Binding site: c

  Query Match             76.2%;  Score 32;  DB 1;  Length 1376;
  Best Local Similarity   57.1%;  Pred. No. 7.7e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db        532 CTCKPNP 538


RESULT 8
MMHUB1
laminin beta-1 chain precursor - human
N;Alternate names: laminin chain B1
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C;Accession: S13547; A28483; A26994; S23566
R;Vuolteenaho, R.; Chow, L.T.; Tryggvason, K.
J. Biol. Chem. 265, 15611-15616, 1990
A;Title: Structure of the human laminin B1 chain gene.
A;Reference number: S13547; MUID:90368768; PMID:1975589
A;Accession: S13547
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1786 <VUO>
A;Cross-references: GB:M61951; GB:J02778; NID:g186911; PIDN:AAA59486.1; PID:g186913
A;Note: the nucleotide sequence was submitted to GenBank, February 1991
R;Pikkarainen, T.; Eddy, R.; Fukushima, Y.; Byers, M.; Shows, T.; Pihlajaniemi, T.; Sara
J. Biol. Chem. 262, 10454-10462, 1987
A;Title: Human laminin B1 chain. A multidomain protein with gene (LAMB1) locus in the q2
A;Reference number: A28483; MUID:87280097; PMID:3611077
A;Accession: A28483
A;Molecule type: mRNA
A;Residues: 1-1786 <PIK>
A;Cross-references: GB:M61951; GB:J02778; NID:g186911; PIDN:AAA59486.1; PID:g186913
R;Jaye, M.; Modi, W.S.; Ricca, G.A.; Mudd, R.; Chiu, I.M.; O'Brien, S.J.; Drohan, W.N.
Am. J. Hum. Genet. 41, 605-615, 1987
A;Title: Isolation of a cDNA clone for the human laminin-B1 chain and its gene localizat
A;Reference number: A26994; MUID:88021029; PMID:3661559
A;Accession: A26994
A;Molecule type: mRNA
A;Residues: 1276-1469,'V',1471-1695,'G',1697-1709 <JAY>
A;Cross-references: EMBL:M20206; NID:g186914; PIDN:AAA59487.1; PID:g186915
R;Vuolteenaho, R.; Kallunki, T.; Chow, L.; Ikonen, J.; Pikkarainen, T.; Tryggvason, K.
in Extracellular Matrix Genes, Sandell L.J. and Boyd C.D., eds., pp. 175-193, Academic P
A;Title: Genes for the human laminin B1 and B2 chains.
A;Reference number: S23566
A;Accession: S23566
A;Molecule type: DNA
A;Residues: 762-1786 <VU2>
A;Note: mRNA was also sequenced
C;Genetics:
A;Gene: GDB:LAMB1
A;Cross-references: GDB:119357; OMIM:150240
A;Map position: 7q31.1-7q31.3
A;Introns: 13/1; 71/3; 117/1; 141/3; 204/3; 226/1; 293/3; 334/1; 397/1; 457/1; 494/3; 52
64/3; 1513/1; 1582/2; 1629/3; 1688/3; 1742/1
C;Complex: Lamininis are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1786/Product: laminin beta-1 chain #status predicted <MAT>
F;22-270/Domain: VI <DOM6>
F;271-548/Domain: V <DOM5>
F;271-332/Domain: laminin-type EGF-like homology <LE01>
F;335-395/Domain: laminin-type EGF-like homology <LE02>
F;398-455/Domain: laminin-type EGF-like homology <LE03>
F;458-507/Domain: laminin-type EGF-like homology <LE04>
F;463-468/Region: cell adhesion #status predicted
F;510-540/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;549-774/Domain: IV <DOM4>
F;662-668/Region: cell adhesion #status predicted
F;773-818/Domain: laminin-type EGF-like homology <LE06>
F;775-1178/Domain: III <DOM3>
F;821-864/Domain: laminin-type EGF-like homology <LE07>
F;867-914/Domain: laminin-type EGF-like homology <LE08>
F;917-973/Domain: laminin-type EGF-like homology <LE09>
F;923-927/Region: cell adhesion #status predicted
F;950-954/Region: cell adhesion #status predicted
F;976-1025/Domain: laminin-type EGF-like homology <LE10>
F;1028-1081/Domain: laminin-type EGF-like homology <LE11>
F;1084-1129/Domain: laminin-type EGF-like homology <LE12>
F;1132-1176/Domain: laminin-type EGF-like homology <LE13>
F;1179-1397/Domain: II <DOM2>
F;1179-1397/Region: heptad repeats
F;1398-1430/Domain: alpha <ALP>
F;1431-1786/Domain: I <DOM1>
F;1431-1786/Region: heptad repeats
F;30-35/Disulfide bonds: #status predicted
F;120,356,519,677,965,1041,1195,1279,1336,1343,1487,1542,1643/Binding site: carbohydrate
F;1179,1182,1785/Disulfide bonds: interchain #status predicted

  Query Match             76.2%;  Score 32;  DB 1;  Length 1786;
  Best Local Similarity   66.7%;  Pred. No. 9.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        529 CSCRPH 534


RESULT 9
MMMSB1
laminin beta-1 chain precursor - mouse
N;Alternate names: laminin chain B1
C;Species: Mus musculus (house mouse)
C;Date: 28-Feb-1986 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C;Accession: A26413; S02679; S05326; S14877; A02871; S02036; S13543
R;Sasaki, M.; Kato, S.; Kohno, K.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 84, 935-939, 1987
A;Title: Sequence of the cDNA encoding the laminin B1 chain reveals a multidomain protei
A;Reference number: A26413; MUID:87147212; PMID:3493487
A;Accession: A26413
A;Molecule type: mRNA
A;Residues: 1-1786 <SAS>
A;Cross-references: EMBL:M15525; NID:g198700
A;Note: translation in GenBank has additional 48 residues at the amino end
R;Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
Biochem. J. 252, 453-461, 1988
A;Title: Structure and distribution of N-linked oligosaccharide chains on various domain
A;Reference number: S02678; MUID:88326259; PMID:2458101
A;Accession: S02679
A;Molecule type: protein
A;Residues: 28-42;932-946 <FUJ>
R;Hartl, L.; Oberbaeumer, I.; Deutzmann, R.
Eur. J. Biochem. 173, 629-635, 1988
A;Title: The N terminus of laminin A chain is homologous to the B chains.
A;Reference number: S00624; MUID:88225080; PMID:3267223
A;Accession: S05326
A;Molecule type: protein
A;Residues: 457-466;854-868;932-946 <HAR>
```

R;Mann, K.; Deutzmann, R.; Timpl, R.
Eur. J. Biochem. 178, 71-80, 1988
A;Title: Characterization of proteolytic fragments of the laminin-nidogen complex and th
A;Reference number: S08895; MUID:89078415; PMID:2462498
A;Accession: S14877
A;Molecule type: protein
A;Residues: 590-620 <MAN>
R;Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.
EMBO J. 3, 2355-2362, 1984
A;Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil a
A;Reference number: A02870; MUID:85051302; PMID:6209134
A;Accession: A02871
A;Molecule type: mRNA
A;Residues: 1292-1530,'MEMP',1535-1691,'C',1693-1748,'N',1750-1786 <BAR>
A;Cross-references: EMBL:X05212; NID:g52861; PIDN:CAA28839.1; PID:g809042
R;Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaeumer, I.; Hartl, L.
Eur. J. Biochem. 177, 35-45, 1988
A;Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-te
A;Reference number: S01790; MUID:89030693; PMID:3181157
A;Accession: S02036
A;Molecule type: protein
A;Residues: 1561-1587 <DEU>
R;Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.
EMBO J. 4, 309-316, 1985
A;Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.
A;Reference number: S13543; MUID:85257455; PMID:3848400
A;Accession: S13543
A;Molecule type: protein
A;Residues: 1700-1748,'N',1750-1759 <PAU>
C;Genetics:
A;Gene: Lamb-1
A;Map position: 12
C;Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C;Function:
A;Description: interact with cells and with other basement membrane proteins to promote
C;Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C;Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1786/Product: laminin beta-1 chain #status predicted <MAT>
F;22-270/Domain: VI <DOM6>
F;271-540/Domain: V <DOM5>
F;271-332/Domain: laminin-type EGF-like homology <LE01>
F;335-395/Domain: laminin-type EGF-like homology <LE02>
F;398-455/Domain: laminin-type EGF-like homology <LE03>
F;458-507/Domain: laminin-type EGF-like homology <LE04>
F;510-540/Domain: laminin-type EGF-like homology #status atypical <LE05>
F;541-772/Domain: IV <DOM4>
F;773-1182/Domain: III <DOM3>
F;773-818/Domain: laminin-type EGF-like homology <LE06>
F;821-864/Domain: laminin-type EGF-like homology <LE07>
F;867-914/Domain: laminin-type EGF-like homology <LE08>
F;917-973/Domain: laminin-type EGF-like homology <LE09>
F;976-1025/Domain: laminin-type EGF-like homology <LE10>
F;1028-1081/Domain: laminin-type EGF-like homology <LE11>
F;1084-1129/Domain: laminin-type EGF-like homology <LE12>
F;1132-1176/Domain: laminin-type EGF-like homology <LE13>
F;1183-1397/Domain: II <DOM2>
F;1183-1397/Region: heptad repeats
F;1398-1430/Domain: alpha <ALP>
F;1431-1786/Region: heptad repeats
F;1431-1786/Domain: I <DOM1>
F;22/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;30-35/Disulfide bonds: #status predicted
F;120,356,519,677,1041,1195,1279,1336,1343,1487,1533,1542,1643/Binding site: carbohydrat
F;1179,1182,1785/Disulfide bonds: interchain #status predicted

  Query Match 76.2%; Score 32; DB 1; Length 1786;
  Best Local Similarity 66.7%; Pred. No. 9.4e+02;
  Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPH 6
| | ||
Db 529 CSCLPH 534

RESULT 10
A48523
margatoxin [validated] - scorpion (Centruroides margaritatus)
N;Alternate names: potassium channel inhibitor MgTx
C;Species: Centruroides margaritatus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 15-Sep-2000
C;Accession: A48523
R;Garcia-Calvo, M.; Leonard, R.J.; Novick, J.; Stevens, S.P.; Schmalhofer, W.; Kaczorowsk
J. Biol. Chem. 268, 18866-18874, 1993
A;Title: Purification, characterization, and biosynthesis of margatoxin, a component of C
A;Reference number: A48523; MUID:93366802; PMID:8360176
A;Accession: A48523
A;Molecule type: protein
A;Residues: 1-39 <GAR>
R;Johnson, B.A.; Stevens, S.P.; Williamson, J.M.
submitted to the Brookhaven Protein Data Bank, December 1994
A;Reference number: A66180; PDB:1MTX
A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, residues
R;Johnson, B.A.; Stevens, S.P.; Williamson, J.M.
Biochemistry 33, 15061-15070, 1994
A;Title: Determination of the three-dimensional structure of margatoxin by (1)H, (13)C,
A;Reference number: A56028; MUID:95092763; PMID:7999764
A;Contents: annotation; conformation and disulfide bond assignments by (1)H-, (13)C-, an
C;Superfamily: kaliotoxin
C;Keywords: neurotoxin; potassium channel inhibitor; venom
F;7-29,13-34,17-36/Disulfide bonds: #status experimental

  Query Match 73.8%; Score 31; DB 1; Length 39;
  Best Local Similarity 66.7%; Pred. No. 71;
  Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPH 6
| | ||
Db 34 CKCYPH 39

RESULT 11
JV0044
hypothetical 7.5K protein (fepB 5' region) - Escherichia coli
C;Species: Escherichia coli
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 18-Jun-1993
C;Accession: JV0044
R;Elkins, M.F.; Earhart, C.F.
J. Bacteriol. 171, 5443-5451, 1989
A;Title: Nucleotide sequence and regulation of the Escherichia coli gene for ferrientero
A;Reference number: A91900; MUID:90008779; PMID:2529253
A;Accession: JV0044
A;Molecule type: DNA
A;Residues: 1-68 <ELK>
A;Experimental source: strain K12
C;Genetics:
A;Map position: 14 min

  Query Match 73.8%; Score 31; DB 2; Length 68;
  Best Local Similarity 80.0%; Pred. No. 1.1e+02;
  Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CXPHP 7
| |||
Db 35 CSPHP 39

RESULT 12
A35666
transcription activator Krox-24 88K - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 21-Sep-1990 #sequence_revision 21-Sep-1990 #text_change 02-Jul-1998
C;Accession: A35666
R;Lemaire, P.; Vesque, C.; Schmitt, J.; Stunnenberg, H.; Frank, R.; Charnay, P.
Mol. Cell. Biol. 10, 3456-3467, 1990
A;Title: The serum-inducible mouse gene Krox-24 encodes a sequence-specific transcriptio

A;Reference number: A35666; MUID:90287135; PMID:2113174
A;Accession: A35666
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-87 <LEM>
A;Cross-references: GB:M38174

Query Match 73.8%; Score 31; DB 2; Length 87;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CXCXPH 6
| | ||
Db 18 CPCTPH 23

RESULT 13
CUPSAM
amicyanin precursor - Methylobacterium extorquens (strain AM1)
C;Species: Methylobacterium extorquens
C;Date: 04-Dec-1986 #sequence_revision 21-Jul-1995 #text_change 20-Apr-2000
C;Accession: A56621; A00295
R;Chistoserdov, A.Y.; Tsygankov, Y.D.; Lidstrom, M.E.
DNA Seq. 2, 53-55, 1991
A;Title: Nucleotide sequence of the amicyanin gene from Methylobacterium extorquens AM1.
A;Reference number: A56621; MUID:92199244; PMID:1802036
A;Accession: A56621
A;Molecule type: DNA
A;Residues: 1-119 <CHI>
A;Cross-references: GB:M57963; NID:g150014; PIDN:AAA68895.1; PID:g150016
A;Note: sequence modified after extraction from NCBI backbone
A;Note: the authors translated the codon CAC for residue 70 as Asn
A;Note: sequence extracted from NCBI backbone (NCBIN:89409, NCBIP:89412)
R;Ambler, R.P.; Tobari, J.
Biochem. J. 232, 451-457, 1985
A;Title: The primary structures of Pseudomonas AM1 amicyanin and pseudoazurin. Two new s
A;Reference number: A90327; MUID:86130354; PMID:4091802
A;Accession: A00295
A;Molecule type: protein
A;Residues: 21-119 <AMB>
C;Comment: This species of Pseudomonas, isolated as an airborne contaminant, uses compou
the true pseudomonads as well as methylotrophs.
C;Superfamily: plastocyanin
C;Keywords: copper; electron transfer; metalloprotein; periplasmic space
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-119/Product: amicyanin #status experimental <MAT>
F;67,106,109,112/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 73.8%; Score 31; DB 1; Length 119;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CXPHP 7
| |||
Db 106 CTPHP 110

RESULT 14
S12972
amicyanin - Paracoccus denitrificans
C;Species: Paracoccus denitrificans
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 20-Apr-2000
C;Accession: S12972
R;van Spanning, R.J.M.; Wansell, C.W.; Reijnders, W.N.M.; Oltmann, L.F.; Stouthamer, A.H
FEBS Lett. 275, 217-220, 1990
A;Title: Mutagenesis of the gene encoding amicyanin of Paracoccus denitrificans and the
A;Reference number: S12971; MUID:91085564; PMID:2261991
A;Accession: S12972
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-131 <SPA>
A;Cross-references: EMBL:X55665; NID:g45458; PIDN:CAA39199.1; PID:g45460
C;Superfamily: plastocyanin
C;Keywords: copper; electron transfer; metalloprotein
F;79,118,121,124/Binding site: copper (His, Cys, His, Met) (type 1) #status experimental

Query Match 73.8%; Score 31; DB 1; Length 131;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CXPHP 7
| |||
Db 118 CTPHP 122

RESULT 15
A23706
amicyanin - Thiobacillus versutus
N;Alternate names: amine dehydrogenase amicyanin component
C;Species: Thiobacillus versutus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S19732; A23706
R;Ubbink, M.; van Kleef, M.A.G.; Kleinjan, D.J.; Hoitink, C.W.G.; Huitema, F.; Beintema,
Eur. J. Biochem. 202, 1003-1012, 1991
A;Title: Cloning, sequencing and expression studies of the genes encoding amicyanin and
A;Reference number: S19730; MUID:92111471; PMID:1765062
A;Accession: S19732
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-132 <UBB>
A;Cross-references: GB:M58001; NID:g154632; PIDN:AAA50571.1; PID:g154635
R;Van Beeumen, J.; Van Bun, S.; Canters, G.W.; Lommen, A.; Chothia, C.
J. Biol. Chem. 266, 4869-4877, 1991
A;Title: The structural homology of amicyanin from Thiobacillus versutus to plant plastc
A;Reference number: A23706; MUID:91161570; PMID:2002033
A;Accession: A23706
A;Status: preliminary
A;Molecule type: protein
A;Residues: 28-132 <VAN>
C;Superfamily: plastocyanin
C;Keywords: copper; electron transfer; metalloprotein
F;80,119,122,125/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 73.8%; Score 31; DB 1; Length 132;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 CXPHP 7
| |||
Db 119 CTPHP 123

Search completed: April 8, 2004, 11:18:04
Job time : 10.5278 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: April 8, 2004, 11:00:44 ; Search time 5.83333 Seconds
(without alignments)
62.484 Million cell updates/sec

Title: US-09-753-139C-1
Perfect score: 42
Sequence: 1 CXCXPHP 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	32	76.2	830	1	SREC_HUMAN	Q14162 homo sapien
2	32	76.2	833	1	SRC2_MOUSE	P59222 mus musculu
3	32	76.2	870	1	SRC2_HUMAN	Q96gp6 homo sapien
4	32	76.2	1172	1	LMB3_HUMAN	Q13751 homo sapien
5	32	76.2	1376	1	VGL2_CVM4	P22432 murine coro
6	32	76.2	1376	1	VGL2_CVMJC	Q02385 murine coro
7	32	76.2	1428	1	ATRN_MOUSE	Q9wu60 mus musculu
8	32	76.2	1587	1	LMG3_HUMAN	Q9y6n6 homo sapien
9	32	76.2	1786	1	LMB1_HUMAN	P07942 homo sapien
10	32	76.2	1786	1	LMB1_MOUSE	P02469 mus musculu
11	31	73.8	31	1	SCT2_MESTA	P59870 mesobuthus
12	31	73.8	39	1	SCK1_CENLM	P59847 centruroide
13	31	73.8	39	1	SCKM_CENMA	P40755 centruroide
14	31	73.8	119	1	AMCY_METEX	P04172 methylobact
15	31	73.8	131	1	AMCY_PARDE	P22364 paracoccus
16	31	73.8	132	1	AMCY_PARVE	P22365 paracoccus
17	31	73.8	235	1	YHBF_BACSU	P39133 bacillus su
18	31	73.8	242	1	SCNC_THITI	O66188 thiobacillu
19	31	73.8	313	1	GAG_AVISN	P03342 avian splee
20	31	73.8	317	1	VN35_ROTBN	Q03241 bovine rota
21	31	73.8	317	1	VN35_ROTBU	P03538 bovine rota
22	31	73.8	317	1	VN35_ROTHD	Q03240 human rotav
23	31	73.8	317	1	VN35_ROTHW	Q03245 human rotav
24	31	73.8	317	1	VN35_ROTP5	P09366 porcine rot
25	31	73.8	317	1	VN35_ROTS1	P03537 simian 11 r
26	31	73.8	317	1	VN35_ROTSP	Q03242 simian 11 r
27	31	73.8	317	1	VN35_ROTSR	Q03243 simian 11 r
28	31	73.8	326	1	BAR1_CHIPA	P08724 chironomus
29	31	73.8	328	1	FMT_BIFLO	Q8g3h1 bifidobacte
30	31	73.8	359	1	OM40_MOUSE	Q9qya2 mus musculu
31	31	73.8	361	1	OM40_HUMAN	O96008 homo sapien
32	31	73.8	399	1	CT26_MOUSE	Q8ce12 mus musculu
33	31	73.8	459	1	Y976_TREPA	O83941 treponema p
34	31	73.8	502	1	P2BC_HUMAN	P48454 homo sapien
35	31	73.8	506	1	ARSA_MOUSE	P50428 mus musculu
36	31	73.8	508	1	LCK_HUMAN	P06239 homo sapien
37	31	73.8	513	1	P2BC_MOUSE	P48455 mus musculu
38	31	73.8	515	1	P2BB_MOUSE	P48453 mus musculu
39	31	73.8	518	1	P2B_NEUCR	Q05681 neurospora
40	31	73.8	521	1	P2BA_BOVIN	P48452 bos taurus
41	31	73.8	521	1	P2BA_HUMAN	Q08209 homo sapien
42	31	73.8	521	1	P2BA_MOUSE	P20652 mus musculu
43	31	73.8	524	1	P2BB_HUMAN	P16299 homo sapien
44	31	73.8	525	1	P2BB_RAT	P20651 rattus norv
45	31	73.8	530	1	P2B_EMENI	P48457 emericella

ALIGNMENTS

RESULT 1
SREC_HUMAN
ID SREC_HUMAN STANDARD; PRT; 830 AA.
AC Q14162; O43701;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Endothelial cells scavenger receptor precursor (Acetyl LDL receptor)
DE (Scavenger receptor class F member 1).
GN SCARF1 OR SREC OR KIAA0149.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=98058897; PubMed=9395444;
RA Adachi H., Tsujimoto M., Arai H., Inoue K.;
RT "Expression cloning of a novel scavenger receptor from human
RT endothelial cells.";
RL J. Biol. Chem. 272:31217-31220(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22086180; PubMed=11978792;
RA Adachi H., Tsujimoto M.;
RT "Characterization of the human gene encoding the scavenger receptor
RT expressed by endothelial cell and its regulation by a novel
RT transcription factor, endothelial zinc finger protein-2.";
RL J. Biol. Chem. 277:24014-24021(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

```
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Mediates the binding and degradation of acetylated low
CC       density lipoprotein (Ac-LDL). Mediates heterophilic interactions,
CC       suggesting a function as adhesion protein (By similarity).
CC   -!- SUBUNIT: Heterophilic interaction with SREC2 via its extracellular
CC       domain. The heterophilic interaction is suppressed by the presence
CC       of ligand such as Ac-LDL (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Endothelial cells.
CC   -!- SIMILARITY: Contains 6 EGF-like domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D86864; BAA24070.1; -.
DR   EMBL; AB052946; BAC02692.1; -.
DR   EMBL; D63483; BAA09770.1; -.
DR   EMBL; BC039735; AAH39735.1; -.
DR   HSSP; P01180; 2BN2.
DR   Genew; HGNC:16820; SCARF1.
DR   MIM; 607873; -.
DR   GO; GO:0016021; C:integral to membrane; IDA.
DR   GO; GO:0030169; F:low-density lipoprotein binding; IDA.
DR   GO; GO:0004888; F:transmembrane receptor activity; TAS.
DR   GO; GO:0045192; P:low-density lipoprotein catabolism; TAS.
DR   GO; GO:0006898; P:receptor mediated endocytosis; TAS.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR009030; Grow_fac_recep.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 5.
DR   PROSITE; PS00022; EGF_1; 6.
DR   PROSITE; PS01186; EGF_2; 6.
DR   PROSITE; PS50026; EGF_3; 3.
KW   Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW   EGF-like domain; Glycoprotein.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    830       ENDOTHELIAL CELLS SCAVENGER RECEPTOR.
FT   DOMAIN       20    421       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    422    442       POTENTIAL.
FT   DOMAIN      443    830       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       53     87       EGF-LIKE 1.
FT   DOMAIN       95    130       EGF-LIKE 2.
FT   DOMAIN      155    191       EGF-LIKE 3.
FT   DOMAIN      215    249       EGF-LIKE 4.
FT   DOMAIN      302    339       EGF-LIKE 5.
FT   DOMAIN      351    382       EGF-LIKE 6.
FT   DOMAIN      476    620       PRO/SER-RICH.
FT   DOMAIN      622    798       GLY-RICH.
FT   DOMAIN      431    438       POLY-LEU.
FT   DISULFID     57     69       POTENTIAL.
FT   DISULFID     63     75       POTENTIAL.
FT   DISULFID     77     86       POTENTIAL.
FT   DISULFID     99    111       POTENTIAL.
FT   DISULFID    105    118       POTENTIAL.
FT   DISULFID    120    129       POTENTIAL.
FT   DISULFID    159    172       POTENTIAL.
FT   DISULFID    165    179       POTENTIAL.
FT   DISULFID    181    190       POTENTIAL.
FT   DISULFID    219    230       POTENTIAL.
FT   DISULFID    225    237       POTENTIAL.
FT   DISULFID    239    248       POTENTIAL.
FT   DISULFID    306    319       POTENTIAL.
FT   DISULFID    313    326       POTENTIAL.
FT   DISULFID    329    338       POTENTIAL.
FT   DISULFID    355    363       POTENTIAL.
FT   DISULFID    358    370       POTENTIAL.
FT   DISULFID    372    381       POTENTIAL.
FT   CARBOHYD    289    289       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    382    382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    393    393       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    662    662       R -> W (IN REF. 3).
SQ   SEQUENCE   830 AA;  87430 MW;  F560D9E1AA64D779 CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 830;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        133 CACGPH 138


RESULT 2
SRC2_MOUSE
ID   SRC2_MOUSE     STANDARD;      PRT;   833 AA.
AC   P59222;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Scavenger receptor class F member 2 precursor (Scavenger receptor
DE   expressed by endothelial cells 2 protein) (SREC-II).
GN   SCARF2 OR SREC2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22267235; PubMed=12154095;
RA   Ishii J., Adachi H., Aoki J., Koizumi H., Tomita S., Suzuki T.,
RA   Tsujimoto M., Inoue K., Arai H.;
RT   "SREC-II, a new member of the scavenger receptor type F family,
RT   trans-interacts with SREC-I through its extracellular domain.";
RL   J. Biol. Chem. 277:39696-39702(2002).
CC   -!- FUNCTION: Probable adhesion protein, which mediates homophilic and
CC       heterophilic interactions. In contrast to SCARF1, it poorly
CC       mediates the binding and degradation of acetylated low density
CC       lipoprotein (Ac-LDL).
CC   -!- SUBUNIT: Homophilic and heterophilic interaction via its
CC       extracellular domain. Interacts with SCARF1. The heterophilic
CC       interaction with SCARF1, which is stronger than the homophilic
CC       interaction with itself, is suppressed by the presence of SCARF1
CC       ligand such as Ac-LDL.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- SIMILARITY: Contains 7 EGF-like domains.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF522197; AAN45862.1; -.
DR   MGD; MGI:1858430; Scarf2.
DR   GO; GO:0005044; F:scavenger receptor activity; IDA.
DR   GO; GO:0007157; P:heterophilic cell adhesion; IDA.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   PRINTS; PR00011; EGFLAMININ.
```

```
DR   SMART; SM00181; EGF; 8.
DR   SMART; SM00180; EGF_Lam; 6.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 3.
KW   Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW   EGF-like domain; Glycoprotein.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34    833       SCAVENGER RECEPTOR CLASS F MEMBER 2.
FT   DOMAIN       34    433       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    434    454       POTENTIAL.
FT   DOMAIN      455    791       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       68    102       EGF-LIKE 1.
FT   DOMAIN      114    145       EGF-LIKE 2.
FT   DOMAIN      146    174       EGF-LIKE 3.
FT   DOMAIN      175    204       EGF-LIKE 4.
FT   DOMAIN      205    233       EGF-LIKE 5.
FT   DOMAIN      234    262       EGF-LIKE 6.
FT   DOMAIN      364    395       EGF-LIKE 7.
FT   DOMAIN      639    714       PRO-RICH.
FT   DISULFID     72     84       POTENTIAL.
FT   DISULFID     78     90       POTENTIAL.
FT   DISULFID     92    101       POTENTIAL.
FT   DISULFID    118    126       POTENTIAL.
FT   DISULFID    120    133       POTENTIAL.
FT   DISULFID    135    144       POTENTIAL.
FT   DISULFID    148    155       POTENTIAL.
FT   DISULFID    150    162       POTENTIAL.
FT   DISULFID    164    173       POTENTIAL.
FT   DISULFID    177    185       POTENTIAL.
FT   DISULFID    179    192       POTENTIAL.
FT   DISULFID    194    203       POTENTIAL.
FT   DISULFID    207    214       POTENTIAL.
FT   DISULFID    209    221       POTENTIAL.
FT   DISULFID    223    232       POTENTIAL.
FT   DISULFID    236    243       POTENTIAL.
FT   DISULFID    238    250       POTENTIAL.
FT   DISULFID    252    261       POTENTIAL.
FT   DISULFID    368    376       POTENTIAL.
FT   DISULFID    371    383       POTENTIAL.
FT   DISULFID    385    394       POTENTIAL.
FT   CARBOHYD     75     75       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    395    395       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   833 AA;  87871 MW;  51EADEEAACAPF005 CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 833;
  Best Local Similarity   66.7%;  Pred. No. 2.1e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        118 CSCHPH 123


RESULT 3
SRC2_HUMAN
ID   SRC2_HUMAN     STANDARD;      PRT;   870 AA.
AC   Q96GP6; Q9BW74;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Scavenger receptor class F member 2 precursor (Scavenger receptor
DE   expressed by endothelial cells 2 protein) (SREC-II).
GN   SCARF2 OR SREC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX   MEDLINE=22267235; PubMed=12154095;
RA   Ishii J., Adachi H., Aoki J., Koizumi H., Tomita S., Suzuki T.,
RA   Tsujimoto M., Inoue K., Arai H.;
RT   "SREC-II, a new member of the scavenger receptor type F family,
RT   trans-interacts with SREC-I through its extracellular domain.";
RL   J. Biol. Chem. 277:39696-39702(2002).
RN   [2]
RP   SEQUENCE OF 272-870 FROM N.A., AND VARIANTS GLU-777 AND LEU-778.
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Probable adhesion protein, which mediates homophilic and
CC       heterophilic interactions. In contrast to SCARF1, it poorly
CC       mediates the binding and degradation of acetylated low density
CC       lipoprotein (Ac-LDL) (By similarity).
CC   -!- SUBUNIT: Homophilic and heterophilic interaction via its
CC       extracellular domain. Interacts with SCARF1. The heterophilic
CC       interaction with SCARF1, which is stronger than the homophilic
CC       interaction with itself, is suppressed by the presence of SCARF1
CC       ligand such as Ac-LDL (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Predominantly expressed in endothelial cells.
CC       Expressed in heart, placenta, lung, kidney, spleen, small
CC       intestine and ovary.
CC   -!- SIMILARITY: Contains 7 EGF-like domains.
CC   -!- CAUTION: Ref.2 sequences differ from that shown due to
CC       frameshifts in positions 750, 751 and 768.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF522196; AAN45861.1; -.
DR   EMBL; BC000584; AAH00584.1; ALT_FRAME.
DR   EMBL; BC009326; AAH09326.1; ALT_FRAME.
DR   Genew; HGNC:19869; SCARF2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   InterPro; IPR002049; Laminin_EGF.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00181; EGF; 7.
DR   SMART; SM00180; EGF_Lam; 6.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 4.
DR   PROSITE; PS50026; EGF_3; 3.
KW   Cell adhesion; Receptor; Repeat; Signal; Transmembrane;
KW   EGF-like domain; Glycoprotein; Polymorphism.
FT   SIGNAL        1     43       POTENTIAL.
FT   CHAIN        44    870       SCAVENGER RECEPTOR CLASS F MEMBER 2.
FT   DOMAIN       44    441       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    442    462       POTENTIAL.
```

```
FT   DOMAIN      463    830       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       76    110       EGF-LIKE 1.
FT   DOMAIN      122    153       EGF-LIKE 2.
FT   DOMAIN      154    182       EGF-LIKE 3.
FT   DOMAIN      183    212       EGF-LIKE 4.
FT   DOMAIN      213    241       EGF-LIKE 5.
FT   DOMAIN      242    270       EGF-LIKE 6.
FT   DOMAIN      372    403       EGF-LIKE 7.
FT   DOMAIN      652    851       PRO-RICH.
FT   DISULFID     80     92       POTENTIAL.
FT   DISULFID     86     98       POTENTIAL.
FT   DISULFID    100    109       POTENTIAL.
FT   DISULFID    126    134       POTENTIAL.
FT   DISULFID    128    141       POTENTIAL.
FT   DISULFID    143    152       POTENTIAL.
FT   DISULFID    156    163       POTENTIAL.
FT   DISULFID    158    170       POTENTIAL.
FT   DISULFID    172    181       POTENTIAL.
FT   DISULFID    185    193       POTENTIAL.
FT   DISULFID    187    200       POTENTIAL.
FT   DISULFID    202    211       POTENTIAL.
FT   DISULFID    215    222       POTENTIAL.
FT   DISULFID    217    229       POTENTIAL.
FT   DISULFID    231    240       POTENTIAL.
FT   DISULFID    244    251       POTENTIAL.
FT   DISULFID    246    258       POTENTIAL.
FT   DISULFID    260    269       POTENTIAL.
FT   DISULFID    376    384       POTENTIAL.
FT   DISULFID    379    391       POTENTIAL.
FT   DISULFID    393    402       POTENTIAL.
FT   CARBOHYD     83     83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    310    310       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    365    365       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    403    403       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     777    777       D -> E (in dbSNP:759611).
FT                                /FTId=VAR_015148.
FT   VARIANT     778    778       V -> L (in dbSNP:759612).
FT                                /FTId=VAR_015149.
FT   VARIANT     819    819       A -> G (in dbSNP:874100).
FT                                /FTId=VAR_015150.
FT   VARIANT     837    837       A -> G (in dbSNP:874101).
FT                                /FTId=VAR_015151.
FT   CONFLICT    474    478       MISSING (IN REF. 2).
FT   CONFLICT    626    641       ALYARVARREARPARA -> GTRPTTTWITHSTAAS (IN
FT                                REF. 2; AAH00584).
SQ   SEQUENCE   870 AA;  92479 MW;  DCB735A50E6E9D1F CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 870;
  Best Local Similarity   66.7%;  Pred. No. 2.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        126 CSCHPH 131


RESULT 4
LMB3_HUMAN
ID   LMB3_HUMAN     STANDARD;      PRT;  1172 AA.
AC   Q13751; O14947; Q14733; Q9UJK4; Q9UJL1;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin beta-3 chain precursor (Laminin 5 beta 3) (Laminin B1k chain)
DE   (Kalinin B1 chain).
GN   LAMB3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95293372; PubMed=7774918;
RA   Pulkkinen L., Gerecke D.R., Christiano A.M., Wagman D.W.,
RA   Burgeson R.E., Uitto J.;
RT   "Cloning of the beta 3 chain gene (LAMB3) of human laminin 5, a
RT   candidate gene in junctional epidermolysis bullosa.";
RL   Genomics 25:192-198(1995).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 172-190.
RX   MEDLINE=94209274; PubMed=7512558;
RA   Gerecke D.R., Wagman D.W., Champliaud M.F., Burgeson R.E.;
RT   "The complete primary structure for a novel laminin chain, the
RT   laminin B1k chain.";
RL   J. Biol. Chem. 269:11073-11080(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21221101; PubMed=11296269;
RA   Robbins P.B., Lin Q., Goodnough J.B., Tian H., Chen X., Khavari P.A.;
RT   "In vivo restoration of laminin 5 beta 3 expression and function in
RT   junctional epidermolysis bullosa.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:5193-5198(2001).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Epidermis;
RX   MEDLINE=96015057; PubMed=8530036;
RA   Morishima Y., Ariyama T., Yamanishi K., Abe T., Ueda E., Yasuno H.,
RA   Inazawa J.;
RT   "Chromosomal loci of 50 human keratinocyte cDNAs assigned by
RT   fluorescence in situ hybridization.";
RL   Genomics 28:273-279(1995).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Grafham D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   VARIANT JEB LEU-679.
RX   MEDLINE=96055522; PubMed=7550237;
RA   Pulkkinen L., McGrath J.A., Christiano A.M., Uitto J.;
RT   "Detection of sequence variants in the gene encoding the beta 3 chain
RT   of laminin 5 (LAMB3).";
RL   Hum. Mutat. 6:77-84(1995).
RN   [7]
RP   VARIANT GABEB LYS-210.
RX   MEDLINE=99068967; PubMed=9767254;
RA   Mellerio J.E., Eady R.A.J., Atherton D.J., Lake B.D., McGrath J.A.;
RT   "E210K mutation in the gene encoding the beta3 chain of laminin-5
RT   (LAMB3) is predictive of a phenotype of generalized atrophic benign
RT   epidermolysis bullosa.";
RL   Br. J. Dermatol. 139:325-331(1998).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The beta-3 chain is a subunit of laminin-5
CC       (epiligrin/kalinin/nicein).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domain VI is globular.
CC   -!- DISEASE: Defects in LAMB3 are a cause of junctional epidermolysis
CC       bullosa gravis (JEB) [MIM:226700]; also known as junctional
CC       epidermolysis bullosa Herlitz-Pearson type. JEB is a blistering
CC       disorder in skin that is characterized by a separation of basal
CC       cells from the basement membrane due to a decreased number of
CC       hemidesmosomes. Laminin-5 is missing from the basement membrane of
CC       patients with the gravis form of epidermolysis bullosa.
CC   -!- DISEASE: Defects in LAMB3 are a cause of generalized atrophic
CC       benign epidermolysis bullosa (GABEB) [MIM:226650]. This nonlethal
CC       form of junctional epidermolysis bullosa is characterized by life-
```

```
CC       long blistering of the skin, associated with hair and tooth
CC       abnormalities.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 6 laminin EGF-like domains.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U17760; AAC51352.1; -.
DR   EMBL; U17745; AAC51352.1; JOINED.
DR   EMBL; U17746; AAC51352.1; JOINED.
DR   EMBL; U17747; AAC51352.1; JOINED.
DR   EMBL; U17748; AAC51352.1; JOINED.
DR   EMBL; U17749; AAC51352.1; JOINED.
DR   EMBL; U17750; AAC51352.1; JOINED.
DR   EMBL; U17751; AAC51352.1; JOINED.
DR   EMBL; U17752; AAC51352.1; JOINED.
DR   EMBL; U17753; AAC51352.1; JOINED.
DR   EMBL; U17754; AAC51352.1; JOINED.
DR   EMBL; U17755; AAC51352.1; JOINED.
DR   EMBL; U17756; AAC51352.1; JOINED.
DR   EMBL; U17757; AAC51352.1; JOINED.
DR   EMBL; U17758; AAC51352.1; JOINED.
DR   EMBL; U17759; AAC51352.1; JOINED.
DR   EMBL; L25541; AAA61834.1; -.
DR   EMBL; AY035783; AAK61364.1; -.
DR   EMBL; D37766; BAA22263.1; -.
DR   EMBL; AL031316; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL023754; CAA19297.1; -.
DR   PIR; A53612; A53612.
DR   HSSP; P02468; 1KLO.
DR   Genew; HGNC:6490; LAMB3.
DR   MIM; 150310; -.
DR   MIM; 226650; -.
DR   MIM; 226700; -.
DR   GO; GO:0008544; P:epidermal differentiation; TAS.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00053; laminin_EGF; 6.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 5.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal;
KW   Disease mutation; Epidermolysis bullosa.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN       18   1172       LAMININ BETA-3 CHAIN.
FT   DOMAIN      18    249       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN     250    315       LAMININ EGF-LIKE 1.
FT   DOMAIN     316    378       LAMININ EGF-LIKE 2.
FT   DOMAIN     379    430       LAMININ EGF-LIKE 3.
FT   DOMAIN     431    480       LAMININ EGF-LIKE 4.
FT   DOMAIN     481    533       LAMININ EGF-LIKE 5.
FT   DOMAIN     534    578       LAMININ EGF-LIKE 6.
FT   DOMAIN     579    785       DOMAIN II.
FT   DOMAIN     786    816       DOMAIN ALPHA.
FT   DOMAIN     817   1170       DOMAIN I.
FT   DOMAIN     723    757       COILED COIL (POTENTIAL).
FT   DOMAIN     831    884       COILED COIL (POTENTIAL).
FT   DOMAIN     948   1133       COILED COIL (POTENTIAL).
FT   DISULFID   250    259       BY SIMILARITY.
FT   DISULFID   252    279       BY SIMILARITY.
FT   DISULFID   281    290       BY SIMILARITY.
FT   DISULFID   293    313       BY SIMILARITY.
FT   DISULFID   316    325       BY SIMILARITY.
FT   DISULFID   318    343       BY SIMILARITY.
FT   DISULFID   346    355       BY SIMILARITY.
FT   DISULFID   358    376       BY SIMILARITY.
FT   DISULFID   379    392       BY SIMILARITY.
FT   DISULFID   381    399       BY SIMILARITY.
FT   DISULFID   401    410       BY SIMILARITY.
FT   DISULFID   413    428       BY SIMILARITY.
FT   DISULFID   431    444       BY SIMILARITY.
FT   DISULFID   433    451       BY SIMILARITY.
FT   DISULFID   453    462       BY SIMILARITY.
FT   DISULFID   465    478       BY SIMILARITY.
FT   DISULFID   481    493       BY SIMILARITY.
FT   DISULFID   483    500       BY SIMILARITY.
FT   DISULFID   502    511       BY SIMILARITY.
FT   DISULFID   519    531       BY SIMILARITY.
FT   DISULFID   534    546       BY SIMILARITY.
FT   DISULFID   536    553       BY SIMILARITY.
FT   DISULFID   555    564       BY SIMILARITY.
FT   DISULFID   567    578       BY SIMILARITY.
FT   DISULFID   581    581       INTERCHAIN (PROBABLE).
FT   DISULFID   584    584       INTERCHAIN (PROBABLE).
FT   DISULFID  1171   1171       INTERCHAIN (PROBABLE).
FT   CARBOHYD   220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   604    604       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   810    810       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT    210    210       E -> K (IN GABEB).
FT                               /FTId=VAR_004170.
FT   VARIANT    679    679       P -> L (IN JEB).
FT                               /FTId=VAR_004171.
FT   CONFLICT   124    124       Q -> R (IN REF. 2).
FT   CONFLICT   269    269       MISSING (IN REF. 2).
FT   CONFLICT   388    388       P -> A (IN REF. 2).
FT   CONFLICT   426    427       QG -> RR (IN REF. 2).
FT   CONFLICT   440    441       RD -> E (IN REF. 2).
FT   CONFLICT   489    500       LSPQCNQFTGQC -> PQPTVQPVHRAV (IN REF. 4).
FT   CONFLICT   603    603       R -> P (IN REF. 2).
FT   CONFLICT   815    815       G -> A (IN REF. 2).

  Query Match            76.2%;  Score 32;  DB 1;  Length 1172;
  Best Local Similarity  66.7%;  Pred. No. 2.9e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        481 CACDPH 486


RESULT 5
VGL2_CVM4
ID   VGL2_CVM4      STANDARD;      PRT;  1376 AA.
AC   P22432;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE   [Contains: Spike protein S1 (90B); Spike protein S2 (90A)].
GN   S.
OS   Murine coronavirus (strain 4) (MHV-4) (Murine hepatitis virus).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=12760;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90085815; PubMed=2556846;
RA   Parker S.E., Gallagher T.M., Buchmeier M.J.;
RT   "Sequence analysis reveals extensive polymorphism and evidence of
RT   deletions within the E2 glycoprotein gene of several strains of
RT   murine hepatitis virus.";
RL   Virology 173:664-673(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=91353381; PubMed=1966429;
RA   Parker S.E., Buchmeier M.J.;
RT   "RNA sequence analysis of the E2 genes of wildtype and
RT   neuroattenuated mutants of MHV-4 reveals a hypervariable domain.";
RL   Adv. Exp. Med. Biol. 276:395-402(1990).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC       AND IN SYNCYTIUM FORMATION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: NEARLY IDENTICAL TO THE E2 GLYCOPROTEINS FROM MVH-JHM
CC       AND MVH-A59 STRAINS, EXCEPT FOR AN N-TERMINAL INSERTION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M32789; AAA46456.1; -.
DR   EMBL; S51114; AAB19590.1; -.
DR   PIR; A33748; VGIHJ2.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     14
FT   CHAIN        15   1376       E2 GLYCOPROTEIN.
FT   CHAIN        15    769       SPIKE PROTEIN S1.
FT   CHAIN       770   1376       SPIKE PROTEIN S2.
FT   DOMAIN       15   1320       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1321   1338       POTENTIAL.
FT   DOMAIN     1339   1376       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1339   1356       CYS-RICH.
FT   DOMAIN      429    599       IMPORTANT FOR THE NEUROVIRULENCE.
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    192    192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    582    582       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    709    709       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    717    717       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    740    740       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    789    789       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    806    806       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    945    945       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1232   1232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1242   1242       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1261   1261       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1277   1277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1298   1298       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1376 AA;  151882 MW;  88C01B97B252094E CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 1376;
  Best Local Similarity   57.1%;  Pred. No. 3.3e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db        532 CTCKPNP 538


RESULT 6
VGL2_CVMJC
ID   VGL2_CVMJC     STANDARD;      PRT;  1376 AA.
AC   Q02385;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   E2 glycoprotein precursor (Spike glycoprotein) (Peplomer protein)
DE   [Contains: Spike protein S1 (90B); Spike protein S2 (90A)].
GN   S.
OS   Murine coronavirus (strain JHMV / variant CL-2) (MHV) (Murine
OS   hepatitis virus).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=33735;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92268864; PubMed=1316938;
RA   Taguchi F., Ikeda T., Shida H.;
RT   "Molecular cloning and expression of a spike protein of neurovirulent
RT   murine coronavirus JHMV variant cl-2.";
RL   J. Gen. Virol. 73:1065-1072(1992).
CC   -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC       TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION
CC       AND IN SYNCYTIUM FORMATION.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10235; BAA01085.1; -.
DR   PIR; JQ1534; JQ1534.
DR   InterPro; IPR002552; Corona_S2.
DR   Pfam; PF01601; Corona_S2; 1.
KW   Glycoprotein; Envelope protein; Transmembrane; Signal.
FT   SIGNAL        1     14
FT   CHAIN        15   1376       E2 GLYCOPROTEIN.
FT   CHAIN        15    769       SPIKE PROTEIN S1.
FT   CHAIN       770   1376       SPIKE PROTEIN S2.
FT   DOMAIN       15   1320       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1321   1338       POTENTIAL.
FT   DOMAIN     1339   1376       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     1339   1356       CYS-RICH.
FT   CARBOHYD     31     31       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     60     60       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    134    134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    192    192       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    357    357       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    435    435       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    709    709       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    717    717       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    789    789       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    806    806       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    945    945       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1232   1232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1242   1242       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1261   1261       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1277   1277       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1298   1298       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1376 AA;  152041 MW;  98C30DDD979F9E75 CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 1376;
  Best Local Similarity   57.1%;  Pred. No. 3.3e+02;
  Matches    4;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPHP 7
              | | |:|
Db        532 CTCKPNP 538


RESULT 7
ATRN_MOUSE
ID   ATRN_MOUSE     STANDARD;      PRT;  1428 AA.
AC   Q9WU60; Q9R263; Q9WU77;
DT   16-OCT-2001 (Rel. 40, Created)
```

```
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Attractin precursor (Mahogany protein).
GN   ATRN OR MGCA OR MG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99184159; PubMed=10086355;
RA   Nagle D.L., McGrail S.H., Vitale J., Woolf E.A., Dussault B.J. Jr.,
RA   DiRocco L., Holmgren L., Montagno J., Bork P., Huszar D.,
RA   Fairchild-Huntress V., Ge P., Keilty J., Ebeling C., Baldini L.,
RA   Gilchrist J., Burn P., Carlson G.A., Moore K.J.;
RT   "The mahogany protein is a receptor involved in suppression of
RT   obesity.";
RL   Nature 398:148-152(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99184160; PubMed=10086356;
RA   Gunn T.M., Miller K.A., He L., Hyman R.W., Davis R.W., Azarani A.,
RA   Schlossman S.F., Duke-Cohan J.S., Barsh G.S.;
RT   "The mouse mahogany locus encodes a transmembrane form of human
RT   attractin.";
RL   Nature 398:152-156(1999).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=20578746; PubMed=11137996;
RA   He L., Gunn T.M., Bouley D.M., Lu X.Y., Watson S.J., Schlossman S.F.,
RA   Duke-Cohan J.S., Barsh G.S.;
RT   "A biochemical function for attractin in agouti-induced pigmentation
RT   and obesity.";
RL   Nat. Genet. 27:40-47(2001).
CC   -!- FUNCTION: May play a role in melanocortin signaling pathways that
CC       regulate energy homeostasis and hair color. Is a low-affinity
CC       receptor for agouti.
CC   -!- FUNCTION: Involved in the initial immune cell clustering during
CC       inflammatory response and may regulate chemotactic activity of
CC       chemokines (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: Contains 1 CUB domain.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: Contains 1 laminin EGF-like domain.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC       frameshift in position 1388.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF116897; AAD20947.1; ALT_FRAME.
DR   EMBL; AF119821; AAD25372.1; -.
DR   EMBL; AF120318; AAD22476.1; -.
DR   EMBL; AF120317; AAD22476.1; JOINED.
DR   HSSP; P02468; 1KLO.
DR   MGD; MGI:1341628; Atrn.
DR   GO; GO:0004872; F:receptor activity; IPI.
DR   GO; GO:0040014; P:regulation of body size; IMP.
DR   InterPro; IPR000859; CUB.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006652; Kelch_rep.
DR   InterPro; IPR002049; LamInn_EGF.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   Pfam; PF00431; CUB; 1.
DR   Pfam; PF01344; Kelch; 5.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF01437; PSI; 4.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00042; CUB; 1.
DR   SMART; SM00180; EGF_Lam; 1.
DR   SMART; SM00423; PSI; 5.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS00022; EGF_1; 3.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   Receptor; Laminin EGF-like domain; EGF-like domain; Glycoprotein;
KW   Transmembrane; Inflammatory response; Lectin; Signal;
KW   Alternative splicing.
FT   SIGNAL        1     17       BY SIMILARITY.
FT   CHAIN        18   1428       ATTRACTIN.
FT   DOMAIN       18   1278       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1279   1299       POTENTIAL.
FT   DOMAIN     1300   1428       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      100    128       EGF-LIKE.
FT   DOMAIN      131    247       CUB.
FT   DOMAIN      794    918       C-TYPE LECTIN.
FT   DOMAIN     1076   1105       LAMININ EGF-LIKE.
FT   DOMAIN       45     76       LEU-RICH.
FT   DOMAIN       60     63       POLY-PRO.
FT   DISULFID    100    110       BY SIMILARITY.
FT   DISULFID    104    117       BY SIMILARITY.
FT   DISULFID    119    128       BY SIMILARITY.
FT   DISULFID   1062   1070       BY SIMILARITY.
FT   DISULFID   1064   1076       BY SIMILARITY.
FT   DISULFID   1079   1088       BY SIMILARITY.
FT   DISULFID   1091   1105       BY SIMILARITY.
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    236    236       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    241    241       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    252    252       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    263    263       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    361    361       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    382    382       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    415    415       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    427    427       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    574    574       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    622    622       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    730    730       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    862    862       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    913    913       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    922    922       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    985    985       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1042   1042       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1053   1053       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1072   1072       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1197   1197       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1205   1205       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1249   1249       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1258   1258       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      6      6       A -> V (IN REF. 2).
FT   CONFLICT      9      9       A -> S (IN REF. 2).
FT   CONFLICT     20     20       T -> A (IN REF. 2).
FT   CONFLICT    164    169       QPNRIM -> YPNAVL (IN REF. 2).
FT   CONFLICT    255    255       G -> A (IN REF. 2).
FT   CONFLICT    505    505       V -> A (IN REF. 2).
FT   CONFLICT    833    833       F -> L (IN REF. 2).
FT   CONFLICT   1075   1075       I -> V (IN REF. 2).
FT   CONFLICT   1140   1140       E -> K (IN REF. 2).
FT   CONFLICT   1171   1171       E -> G (IN REF. 2).
FT   CONFLICT   1192   1192       L -> F (IN REF. 2).
FT   CONFLICT   1201   1201       N -> K (IN REF. 2).
SQ   SEQUENCE   1428 AA;  158087 MW;  D1EADBF7F53B8911 CRC64;
```

```
  Query Match             76.2%;  Score 32;  DB 1;  Length 1428;
  Best Local Similarity   66.7%;  Pred. No. 3.4e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        306 CSCFPH 311


RESULT 8
LMG3_HUMAN
ID   LMG3_HUMAN     STANDARD;      PRT;  1587 AA.
AC   Q9Y6N6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin gamma-3 chain precursor (Laminin 12 gamma 3).
GN   LAMC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=99242614; PubMed=10225960;
RA   Koch M., Olson P.F., Albus A., Jin W., Hunter D.D., Brunken W.J.,
RA   Burgeson R.E., Champliaud M.F.;
RT   "Characterization and expression of the laminin gamma3 chain: a novel,
RT   non-basement membrane-associated, laminin chain.";
RL   J. Cell Biol. 145:605-618(1999).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The gamma-3 chain is a subunit of laminin-12.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Broadly expressed in: skin, heart, lung, and
CC       the reproductive tracts.
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domain IV is globular.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 11 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin IV domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF041835; AAD36991.1; -.
DR   HSSP; P02468; 1TLE.
DR   Genew; HGNC:6494; LAMC3.
DR   MIM; 604349; -.
DR   GO; GO:0005578; C:extracellular matrix; TAS.
DR   GO; GO:0016020; C:membrane; TAS.
DR   GO; GO:0005198; F:structural molecule activity; TAS.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR008212; Lam_N2.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00052; laminin_B; 1.
DR   Pfam; PF00053; laminin_EGF; 9.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   ProDom; PD002082; Lam_N2; 1.
DR   SMART; SM00180; EGF_Lam; 9.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 7.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 10.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1587       LAMININ GAMMA-3 CHAIN.
FT   DOMAIN       20    270       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      271    326       LAMININ EGF-LIKE 1.
FT   DOMAIN      327    382       LAMININ EGF-LIKE 2.
FT   DOMAIN      383    429       LAMININ EGF-LIKE 3.
FT   DOMAIN      430    479       LAMININ EGF-LIKE 4.
FT   DOMAIN      480    489       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN      490    672       LAMININ DOMAIN IV.
FT   DOMAIN      673    706       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN      707    754       LAMININ EGF-LIKE 6.
FT   DOMAIN      755    809       LAMININ EGF-LIKE 7.
FT   DOMAIN      810    865       LAMININ EGF-LIKE 8.
FT   DOMAIN      866    916       LAMININ EGF-LIKE 9.
FT   DOMAIN      917    964       LAMININ EGF-LIKE 10.
FT   DOMAIN      965   1013       LAMININ EGF-LIKE 11.
FT   DOMAIN     1014   1587       DOMAIN II AND I.
FT   DOMAIN     1071   1141       COILED COIL (POTENTIAL).
FT   DOMAIN     1200   1229       COILED COIL (POTENTIAL).
FT   DOMAIN     1424   1504       COILED COIL (POTENTIAL).
FT   DOMAIN     1535   1579       COILED COIL (POTENTIAL).
FT   SITE       1059   1061       CELL ATTACHMENT SITE (POTENTIAL).
FT   CARBOHYD     87     87       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    295    295       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    631    631       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    837    837       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    980    980       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1185   1185       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1518   1518       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1587 AA;  172051 MW;  3CB6E09B5F203319 CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 1587;
  Best Local Similarity   66.7%;  Pred. No. 3.8e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        887 CSCLPH 892


RESULT 9
LMB1_HUMAN
ID   LMB1_HUMAN     STANDARD;      PRT;  1786 AA.
AC   P07942;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin beta-1 chain precursor (Laminin B1 chain).
GN   LAMB1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90368768; PubMed=1975589;
RA   Vuolteenaho R., Chow L.T., Tryggvason K.;
RT   "Structure of the human laminin B1 chain gene.";
RL   J. Biol. Chem. 265:15611-15616(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=87280097; PubMed=3611077;
RA   Pikkarainen T., Eddy R., Fukushima Y., Byers M., Shows T.,
RA   Pihlajaniemi T., Saraste M., Tryggvason K.;
RT   "Human laminin B1 chain. A multidomain protein with gene (LAMB1)
RT   locus in the q22 region of chromosome 7.";
RL   J. Biol. Chem. 262:10454-10462(1987).
RN   [3]
RP   SEQUENCE OF 1276-1709 FROM N.A.
RX   MEDLINE=88021029; PubMed=3661559;
RA   Jaye M., Modi W.S., Ricca G.A., Mudd R., Chiu I.M., O'Brien S.J.,
RA   Drohan W.N.;
RT   "Isolation of a cDNA clone for the human laminin-B1 chain and its
RT   gene localization.";
RL   Am. J. Hum. Genet. 41:605-615(1987).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The beta-1 chain is a subunit of laminin-1 (EHS laminin),
CC       laminin-2 (merosin), and laminin-6 (K-laminin).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- DOMAIN: The alpha-helical domains I and II are thought to interact
CC       with other laminin chains to form a coiled coil structure.
CC   -!- DOMAIN: Domains VI and IV are globular.
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 13 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin IV domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61951; AAA59486.1; -.
DR   EMBL; M58147; AAA59486.1; JOINED.
DR   EMBL; M61917; AAA59486.1; JOINED.
DR   EMBL; M61918; AAA59486.1; JOINED.
DR   EMBL; M61921; AAA59486.1; JOINED.
DR   EMBL; M61922; AAA59486.1; JOINED.
DR   EMBL; M61923; AAA59486.1; JOINED.
DR   EMBL; M61924; AAA59486.1; JOINED.
DR   EMBL; M61925; AAA59486.1; JOINED.
DR   EMBL; M61926; AAA59486.1; JOINED.
DR   EMBL; M61927; AAA59486.1; JOINED.
DR   EMBL; M61928; AAA59486.1; JOINED.
DR   EMBL; M61929; AAA59486.1; JOINED.
DR   EMBL; M61930; AAA59486.1; JOINED.
DR   EMBL; M61931; AAA59486.1; JOINED.
DR   EMBL; M61932; AAA59486.1; JOINED.
DR   EMBL; M61933; AAA59486.1; JOINED.
DR   EMBL; M61934; AAA59486.1; JOINED.
DR   EMBL; M61935; AAA59486.1; JOINED.
DR   EMBL; M61936; AAA59486.1; JOINED.
DR   EMBL; M61938; AAA59486.1; JOINED.
DR   EMBL; M61939; AAA59486.1; JOINED.
DR   EMBL; M61940; AAA59486.1; JOINED.
DR   EMBL; M61941; AAA59486.1; JOINED.
DR   EMBL; M61942; AAA59486.1; JOINED.
DR   EMBL; M61943; AAA59486.1; JOINED.
DR   EMBL; M61944; AAA59486.1; JOINED.
DR   EMBL; M61945; AAA59486.1; JOINED.
DR   EMBL; M61946; AAA59486.1; JOINED.
DR   EMBL; M61947; AAA59486.1; JOINED.
DR   EMBL; M61948; AAA59486.1; JOINED.
DR   EMBL; M61949; AAA59486.1; JOINED.
DR   EMBL; M61950; AAA59486.1; JOINED.
DR   EMBL; M55370; AAA59485.1; -.
DR   EMBL; M55378; AAA59485.1; JOINED.
DR   EMBL; M55365; AAA59485.1; JOINED.
DR   EMBL; M55371; AAA59485.1; JOINED.
DR   EMBL; M55372; AAA59485.1; JOINED.
DR   EMBL; M55373; AAA59485.1; JOINED.
DR   EMBL; M55374; AAA59485.1; JOINED.
DR   EMBL; M55375; AAA59485.1; JOINED.
DR   EMBL; M55376; AAA59485.1; JOINED.
DR   EMBL; M55344; AAA59485.1; JOINED.
DR   EMBL; M55345; AAA59485.1; JOINED.
DR   EMBL; M55346; AAA59485.1; JOINED.
DR   EMBL; M55347; AAA59485.1; JOINED.
DR   EMBL; M55348; AAA59485.1; JOINED.
DR   EMBL; M55349; AAA59485.1; JOINED.
DR   EMBL; M55350; AAA59485.1; JOINED.
DR   EMBL; M55351; AAA59485.1; JOINED.
DR   EMBL; M55352; AAA59485.1; JOINED.
DR   EMBL; M55353; AAA59485.1; JOINED.
DR   EMBL; M55355; AAA59485.1; JOINED.
DR   EMBL; M55356; AAA59485.1; JOINED.
DR   EMBL; M55357; AAA59485.1; JOINED.
DR   EMBL; M55358; AAA59485.1; JOINED.
DR   EMBL; M55359; AAA59485.1; JOINED.
DR   EMBL; M55360; AAA59485.1; JOINED.
DR   EMBL; M55361; AAA59485.1; JOINED.
DR   EMBL; M55362; AAA59485.1; JOINED.
DR   EMBL; M55363; AAA59485.1; JOINED.
DR   EMBL; M55364; AAA59485.1; JOINED.
DR   EMBL; M55366; AAA59485.1; JOINED.
DR   EMBL; M55367; AAA59485.1; JOINED.
DR   EMBL; M55368; AAA59485.1; JOINED.
DR   EMBL; M55369; AAA59485.1; JOINED.
DR   EMBL; M61916; AAA59482.1; -.
DR   EMBL; M20206; AAA59487.1; -.
DR   PIR; S13547; MMHUB1.
DR   HSSP; P02468; 1KLO.
DR   Genew; HGNC:6486; LAMB1.
DR   MIM; 150240; -.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00053; laminin_EGF; 13.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 12.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism.
FT   SIGNAL       1     21
FT   CHAIN       22   1786       LAMININ BETA-1 CHAIN.
FT   DOMAIN      22    270       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN     271    334       LAMININ EGF-LIKE 1.
FT   DOMAIN     335    397       LAMININ EGF-LIKE 2.
FT   DOMAIN     398    457       LAMININ EGF-LIKE 3.
FT   DOMAIN     458    509       LAMININ EGF-LIKE 4.
FT   DOMAIN     510    540       LAMININ EGF-LIKE 5 (INCOMPLETE).
FT   DOMAIN     541    771       LAMININ DOMAIN IV.
FT   DOMAIN     773    820       LAMININ EGF-LIKE 6.
FT   DOMAIN     821    866       LAMININ EGF-LIKE 7.
FT   DOMAIN     867    916       LAMININ EGF-LIKE 8.
FT   DOMAIN     917    975       LAMININ EGF-LIKE 9.
FT   DOMAIN     976   1027       LAMININ EGF-LIKE 10.
FT   DOMAIN    1028   1083       LAMININ EGF-LIKE 11.
FT   DOMAIN    1084   1131       LAMININ EGF-LIKE 12.
FT   DOMAIN    1132   1178       LAMININ EGF-LIKE 13.
FT   DOMAIN    1179   1397       DOMAIN II.
FT   DOMAIN    1398   1430       DOMAIN ALPHA.
```

```
FT   DOMAIN     1431   1786       DOMAIN I.
FT   DOMAIN     1216   1315       COILED COIL (POTENTIAL).
FT   DOMAIN     1353   1388       COILED COIL (POTENTIAL).
FT   DOMAIN     1442   1781       COILED COIL (POTENTIAL).
FT   DISULFID    271    280       BY SIMILARITY.
FT   DISULFID    273    298       BY SIMILARITY.
FT   DISULFID    300    309       BY SIMILARITY.
FT   DISULFID    312    332       BY SIMILARITY.
FT   DISULFID    335    344       BY SIMILARITY.
FT   DISULFID    337    362       BY SIMILARITY.
FT   DISULFID    365    374       BY SIMILARITY.
FT   DISULFID    377    395       BY SIMILARITY.
FT   DISULFID    398    411       BY SIMILARITY.
FT   DISULFID    400    426       BY SIMILARITY.
FT   DISULFID    428    437       BY SIMILARITY.
FT   DISULFID    440    455       BY SIMILARITY.
FT   DISULFID    458    472       BY SIMILARITY.
FT   DISULFID    460    479       BY SIMILARITY.
FT   DISULFID    481    490       BY SIMILARITY.
FT   DISULFID    493    507       BY SIMILARITY.
FT   DISULFID    773    785       BY SIMILARITY.
FT   DISULFID    775    792       BY SIMILARITY.
FT   DISULFID    794    803       BY SIMILARITY.
FT   DISULFID    806    818       BY SIMILARITY.
FT   DISULFID    821    833       BY SIMILARITY.
FT   DISULFID    823    840       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    864       BY SIMILARITY.
FT   DISULFID    867    876       BY SIMILARITY.
FT   DISULFID    869    883       BY SIMILARITY.
FT   DISULFID    886    895       BY SIMILARITY.
FT   DISULFID    898    914       BY SIMILARITY.
FT   DISULFID    917    933       BY SIMILARITY.
FT   DISULFID    919    944       BY SIMILARITY.
FT   DISULFID    946    955       BY SIMILARITY.

  Query Match             76.2%;  Score 32;  DB 1;  Length 1786;
  Best Local Similarity   66.7%;  Pred. No. 4.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        529 CSCRPH 534


RESULT 10
LMB1_MOUSE
ID   LMB1_MOUSE     STANDARD;      PRT;  1786 AA.
AC   P02469;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-JUL-1989 (Rel. 11, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Laminin beta-1 chain precursor (Laminin B1 chain).
GN   LAMB1-1 OR LAMB-1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87147212; PubMed=3493487;
RA   Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;
RT   "Sequence of the cDNA encoding the laminin B1 chain reveals a
RT   multidomain protein containing cysteine-rich repeats.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).
RN   [2]
RP   SEQUENCE OF 1292-1786 FROM N.A.
RX   MEDLINE=85051302; PubMed=6209134;
RA   Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
RT   "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
RT   coiled-coil alpha-helix.";
RL   EMBO J. 3:2355-2362(1984).
RN   [3]
RP   SEQUENCE OF 165-172; 539-547 AND 712-719.
RC   STRAIN=BALB/c; TISSUE=Endothelial cells;
RX   MEDLINE=97363207; PubMed=9219532;
RA   Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,
RA   Sorokin L.M.;
RT   "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of
RT   endothelium.";
RL   Eur. J. Biochem. 246:727-735(1997).
CC   -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC       is thought to mediate the attachment, migration and organization
CC       of cells into tissues during embryonic development by interacting
CC       with other extracellular matrix components.
CC   -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC       different polypeptide chains (alpha, beta, gamma), which are bound
CC       to each other by disulfide bonds into a cross-shaped molecule
CC       comprising one long and three short arms with globules at each
CC       end. The beta-1 chain is a subunit of laminin-1 (EHS laminin),
CC       laminin-2 (merosin), and laminin-6 (K-laminin).
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: Found in the basement membranes (major
CC       component).
CC   -!- SIMILARITY: Contains 1 laminin N-terminal domain.
CC   -!- SIMILARITY: Contains 13 laminin EGF-like domains.
CC   -!- SIMILARITY: Contains 1 laminin IV domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M15525; AAA39407.1; ALT_INIT.
DR   EMBL; X05212; CAA28839.1; -.
DR   PIR; A26413; MMMSB1.
DR   HSSP; P02468; 1KLO.
DR   MGD; MGI:96743; Lamb1-1.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR008211; LamNT.
DR   Pfam; PF00053; laminin_EGF; 13.
DR   Pfam; PF00055; laminin_Nterm; 1.
DR   PRINTS; PR00011; EGFLAMININ.
DR   SMART; SM00180; EGF_Lam; 11.
DR   SMART; SM00136; LamNT; 1.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
KW   Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW   Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22   1786       LAMININ BETA-1 CHAIN.
FT   DOMAIN       22    270       LAMININ N-TERMINAL (DOMAIN VI).
FT   DOMAIN      271    334       LAMININ EGF-LIKE 1.
FT   DOMAIN      335    397       LAMININ EGF-LIKE 2.
FT   DOMAIN      398    457       LAMININ EGF-LIKE 3.
FT   DOMAIN      458    509       LAMININ EGF-LIKE 4.
FT   DOMAIN      510    540       LAMININ EGF-LIKE 5 (INCOMPLETE).
FT   DOMAIN      541    772       LAMININ DOMAIN IV.
FT   DOMAIN      773    820       LAMININ EGF-LIKE 6.
FT   DOMAIN      821    866       LAMININ EGF-LIKE 7.
FT   DOMAIN      867    916       LAMININ EGF-LIKE 8.
FT   DOMAIN      917    975       LAMININ EGF-LIKE 9.
FT   DOMAIN      976   1027       LAMININ EGF-LIKE 10.
FT   DOMAIN     1028   1083       LAMININ EGF-LIKE 11.
FT   DOMAIN     1084   1131       LAMININ EGF-LIKE 12.
FT   DOMAIN     1132   1178       LAMININ EGF-LIKE 13.
FT   DOMAIN     1179   1397       DOMAIN II.
FT   DOMAIN     1398   1430       DOMAIN ALPHA.
FT   DOMAIN     1431   1786       DOMAIN I.
FT   DOMAIN     1216   1315       COILED COIL (POTENTIAL).
FT   DOMAIN     1368   1388       COILED COIL (POTENTIAL).
```

```
FT   DOMAIN     1448   1778       COILED COIL (POTENTIAL).
FT   DISULFID    271    280       BY SIMILARITY.
FT   DISULFID    273    298       BY SIMILARITY.
FT   DISULFID    300    309       BY SIMILARITY.
FT   DISULFID    312    332       BY SIMILARITY.
FT   DISULFID    335    344       BY SIMILARITY.
FT   DISULFID    337    362       BY SIMILARITY.
FT   DISULFID    365    374       BY SIMILARITY.
FT   DISULFID    377    395       BY SIMILARITY.
FT   DISULFID    398    411       BY SIMILARITY.
FT   DISULFID    400    426       BY SIMILARITY.
FT   DISULFID    428    437       BY SIMILARITY.
FT   DISULFID    440    455       BY SIMILARITY.
FT   DISULFID    458    472       BY SIMILARITY.
FT   DISULFID    460    479       BY SIMILARITY.
FT   DISULFID    481    490       BY SIMILARITY.
FT   DISULFID    493    507       BY SIMILARITY.
FT   DISULFID    773    785       BY SIMILARITY.
FT   DISULFID    775    792       BY SIMILARITY.
FT   DISULFID    794    803       BY SIMILARITY.
FT   DISULFID    806    818       BY SIMILARITY.
FT   DISULFID    821    833       BY SIMILARITY.
FT   DISULFID    823    840       BY SIMILARITY.
FT   DISULFID    842    851       BY SIMILARITY.
FT   DISULFID    854    864       BY SIMILARITY.
FT   DISULFID    867    876       BY SIMILARITY.
FT   DISULFID    869    883       BY SIMILARITY.
FT   DISULFID    886    895       BY SIMILARITY.
FT   DISULFID    898    914       BY SIMILARITY.
FT   DISULFID    917    933       BY SIMILARITY.
FT   DISULFID    919    944       BY SIMILARITY.
FT   DISULFID    946    955       BY SIMILARITY.
FT   DISULFID    958    973       BY SIMILARITY.
FT   DISULFID    976    990       BY SIMILARITY.
FT   DISULFID    978    997       BY SIMILARITY.
FT   DISULFID   1000   1009       BY SIMILARITY.
FT   DISULFID   1012   1025       BY SIMILARITY.
FT   DISULFID   1084   1096       BY SIMILARITY.
FT   DISULFID   1086   1103       BY SIMILARITY.
FT   DISULFID   1105   1114       BY SIMILARITY.
FT   DISULFID   1117   1129       BY SIMILARITY.
FT   DISULFID   1132   1144       BY SIMILARITY.
FT   DISULFID   1134   1151       BY SIMILARITY.
FT   DISULFID   1153   1162       BY SIMILARITY.
FT   DISULFID   1165   1176       BY SIMILARITY.
FT   DISULFID   1179   1179       INTERCHAIN (PROBABLE).
FT   DISULFID   1182   1182       INTERCHAIN (PROBABLE).
FT   DISULFID   1785   1785       INTERCHAIN (PROBABLE).
FT   CARBOHYD    120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    677    677       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1041   1041       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1195   1195       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1279   1279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1336   1336       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1343   1343       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1487   1487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1533   1533       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1542   1542       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1643   1643       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   1531   1534       SGNA -> MEMP (IN REF. 2).
FT   CONFLICT   1749   1749       D -> N (IN REF. 2).
SQ   SEQUENCE   1786 AA;  196904 MW;  846671B7BF41A474 CRC64;

  Query Match             76.2%;  Score 32;  DB 1;  Length 1786;
  Best Local Similarity   66.7%;  Pred. No. 4.2e+02;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db        529 CSCLPH 534
```

```
RESULT 11
SCT2_MESTA
ID   SCT2_MESTA     STANDARD;      PRT;    31 AA.
AC   P59870;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Tamapin-2.
OS   Mesobuthus tamulus (Eastern Indian scorpion) (Buthus tamulus).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC   Buthoidea; Buthidae; Mesobuthus.
OX   NCBI_TaxID=34647;
RN   [1]
RP   SEQUENCE, MASS SPECTROMETRY, AND FUNCTION.
RC   TISSUE=Venom;
RX   MEDLINE=22336377; PubMed=12239213;
RA   Pedarzani P., D'hoedt D., Doorty K.B., Wadsworth J.D.F., Joseph J.S.,
RA   Jeyaseelan K., Kini R.M., Gadre S.V., Sapatnekar S.M., Stocker M.,
RA   Strong P.N.;
RT   "Tamapin, a venom peptide from the Indian red scorpion (Mesobuthus
RT   tamulus) that targets small conductance Ca2+-activated K+ channels
RT   and afterhyperpolarization currents in central neurons.";
RL   J. Biol. Chem. 277:46101-46109(2002).
CC   -!- FUNCTION: Blocks small conductance calcium-activated potassium
CC       channels.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC   -!- MASS SPECTROMETRY: MW=3431.4; MW_ERR=0.2; METHOD=Electrospray.
CC   -!- SIMILARITY: Belongs to the short scorpion toxin family. Potassium
CC       channel inhibitors subfamily.
DR   PROSITE; PS01138; SCORP_SHORT_TOXIN; 1.
KW   Toxin; Neurotoxin; Ionic channel inhibitor;
KW   Potassium channel inhibitor; Amidation.
FT   DISULFID      3     21       BY SIMILARITY.
FT   DISULFID      8     26       BY SIMILARITY.
FT   DISULFID     12     28       BY SIMILARITY.
FT   MOD_RES      31     31       AMIDATION (PROBABLE).
SQ   SEQUENCE   31 AA;  3439 MW;  E4D43CDC54CA415C CRC64;

  Query Match             73.8%;  Score 31;  DB 1;  Length 31;
  Best Local Similarity   66.7%;  Pred. No. 15;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         26 CKCVPH 31


RESULT 12
SCK1_CENLM
ID   SCK1_CENLM     STANDARD;      PRT;    39 AA.
AC   P59847;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hongotoxin 1 (HgTX1).
OS   Centruroides limbatus (Scorpion).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC   Buthoidea; Buthidae; Centruroides.
OX   NCBI_TaxID=244936;
RN   [1]
RP   SEQUENCE, MUTAGENESIS, AND PHARMACOLOGICAL CHARACTERIZATION.
RC   TISSUE=Venom;
RX   MEDLINE=98112806; PubMed=9446567;
RA   Koschak A., Bugianesi R.M., Mitterdorfer J., Kaczorowski G.J.,
RA   Garcia M.L., Knaus H.-G.;
RT   "Subunit composition of brain voltage-gated potassium channels
RT   determined by hongotoxin-1, a novel peptide derived from Centruroides
RT   limbatus venom.";
RL   J. Biol. Chem. 273:2639-2644(1998).
RN   [2]
RP   STRUCTURE BY NMR, MASS SPECTROMETRY, AND MUTAGENESIS.
```

```
RC   TISSUE=Venom;
RX   MEDLINE=22005852; PubMed=12009929;
RA   Pragl B., Koschak A., Trieb M., Obermair G., Kaufmann W.A.,
RA   Gerster U., Blanc E., Hahn C., Prinz H., Schuetz G., Darbon H.,
RA   Gruber H.J., Knaus H.-G.;
RT   "Synthesis, characterization, and application of cy-dye- and
RT   alexa-dye-labeled hongotoxin(1) analogues. The first high affinity
RT   fluorescence probes for voltage-gated K+ channels.";
RL   Bioconj. Chem. 13:416-425(2002).
CC   -!- FUNCTION: Potent selective inhibitor of Kv1.1, Kv1.2, Kv1.3
CC       voltage-gated potassium channels. Weak inhibitor of Kv1.6
CC       potassium channel. Does not block Kv1.4 and Kv1.5 currents.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC   -!- MASS SPECTROMETRY: MW=4219; METHOD=Electrospray.
CC   -!- SIMILARITY: Belongs to the short scorpion toxin family. Potassium
CC       channel inhibitor subfamily.
DR   PDB; 1HLY; PRELIMINARY.
DR   ProDom; PD003586; Scorpion_toxinS; 1.
DR   PROSITE; PS01138; SCORP_SHORT_TOXIN; 1.
KW   Toxin; Neurotoxin; Ionic channel inhibitor;
KW   Potassium channel inhibitor; 3D-structure.
FT   MUTAGEN      19     19       A->C: NO LOSS OF ACTIVITY.
FT   MUTAGEN      19     19       A->Y: NO LOSS OF ACTIVITY WHEN ASSOCIATED
FT                                WITH F-37.
FT   MUTAGEN      37     37       Y->F: NO LOSS OF ACTIVITY WHEN ASSOCIATED
FT                                WITH Y-19.
SQ   SEQUENCE   39 AA;  4226 MW;  B4ABC83F8EE7E637 CRC64;

  Query Match             73.8%;  Score 31;  DB 1;  Length 39;
  Best Local Similarity   66.7%;  Pred. No. 19;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         34 CKCYPH 39


RESULT 13
SCKM_CENMA
ID   SCKM_CENMA     STANDARD;      PRT;    39 AA.
AC   P40755;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Margatoxin (MgTX).
OS   Centruroides margaritatus (Scorpion).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC   Buthoidea; Buthidae; Centruroides.
OX   NCBI_TaxID=29018;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=93366802; PubMed=8360176;
RA   Garcia-Calvo M., Leonard R.J., Novick J., Stevens S.P.,
RA   Schmalhofer W., Kaczorowski G.J., Garcia M.L.;
RT   "Purification, characterization, and biosynthesis of margatoxin, a
RT   component of Centruroides margaritatus venom that selectively
RT   inhibits voltage-dependent potassium channels.";
RL   J. Biol. Chem. 268:18866-18874(1993).
RN   [2]
RP   SYNTHESIS, AND DISULFIDE BONDS.
RX   MEDLINE=94128107; PubMed=8297371;
RA   Bednarek M.A., Bugianesi R.M., Leonard R.J., Felix J.P.;
RT   "Chemical synthesis and structure-function studies of margatoxin, a
RT   potent inhibitor of voltage-dependent potassium channel in human T
RT   lymphocytes.";
RL   Biochem. Biophys. Res. Commun. 198:619-625(1994).
RN   [3]
RP   STRUCTURE BY NMR.
RX   MEDLINE=95092763; PubMed=7999764;
RA   Johnson B.A., Stevens S.P., Williamson J.M.;
RT   "Determination of the three-dimensional structure of margatoxin by
RT   1H, 13C, 15N triple-resonance nuclear magnetic resonance
RT   spectroscopy.";
RL   Biochemistry 33:15061-15070(1994).
CC   -!- FUNCTION: Potent selective inhibitor of voltage-dependent
CC       potassium channels such as Kv1.3.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Expressed by the venom gland.
CC   -!- SIMILARITY: Belongs to the short scorpion toxin family. Potassium
CC       channel inhibitor subfamily.
DR   PIR; A48523; A48523.
DR   PDB; 1MTX; 14-NOV-95.
DR   InterPro; IPR001947; Scorpion_toxinS.
DR   Pfam; PF00451; toxin_2; 1.
DR   ProDom; PD003586; Scorpion_toxinS; 1.
DR   PROSITE; PS01138; SCORP_SHORT_TOXIN; 1.
KW   Toxin; Neurotoxin; Ionic channel inhibitor;
KW   Potassium channel inhibitor; 3D-structure.
FT   DISULFID      7     29
FT   DISULFID     13     34
FT   DISULFID     17     36
FT   SITE         37     39       INTERACTION WITH KV 1.3 CHANNELS
FT                                (POTENTIAL).
FT   STRAND        6      6
FT   HELIX        10     13
FT   HELIX        14     20
FT   TURN         21     21
FT   TURN         23     24
FT   STRAND       26     30
FT   TURN         31     32
FT   STRAND       33     37
SQ   SEQUENCE   39 AA;  4185 MW;  0FB748318014BE0D CRC64;

  Query Match             73.8%;  Score 31;  DB 1;  Length 39;
  Best Local Similarity   66.7%;  Pred. No. 19;
  Matches    4;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 CXCXPH 6
              | | ||
Db         34 CKCYPH 39


RESULT 14
AMCY_METEX
ID   AMCY_METEX     STANDARD;      PRT;   119 AA.
AC   P04172;
DT   20-MAR-1987 (Rel. 04, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Amicyanin-alpha precursor.
GN   MAUC.
OS   Methylobacterium extorquens.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Methylobacteriaceae; Methylobacterium.
OX   NCBI_TaxID=408;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AM1 / NCIMB 9133;
RX   MEDLINE=91358385; PubMed=1653226;
RA   Chistoserdov A.Y., Tsygankov Y.D., Lidstrom M.E.;
RT   "Genetic organization of methylamine utilization genes from
RT   Methylobacterium extorquens AM1.";
RL   J. Bacteriol. 173:5901-5908(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AM1 / NCIMB 9133;
RX   MEDLINE=94292425; PubMed=8021187;
RA   Chistoserdov A.Y., Chistoserdova L.V., McIntire W.S., Lidstrom M.E.;
RT   "Genetic organization of the mau gene cluster in Methylobacterium
RT   extorquens AM1: complete nucleotide sequence and generation and
RT   characteristics of mau mutants.";
RL   J. Bacteriol. 176:4052-4065(1994).
RN   [3]
RP   SEQUENCE OF 21-119.
```

```
RC   STRAIN=AM1 / NCIMB 9133;
RX   MEDLINE=86130354; PubMed=4091802;
RA   Ambler R.P., Tobari J.;
RT   "The primary structures of Pseudomonas AM1 amicyanin and
RT   pseudoazurin. Two new sequence classes of blue copper proteins.";
RL   Biochem. J. 232:451-457(1985).
CC   -!- FUNCTION: Primary acceptor of electrons from methylamine
CC       dehydrogenase. Passes those electrons on either a soluble
CC       cytochrome c or to pseudoazurin.
CC   -!- COFACTOR: Binds 1 copper ion per molecule.
CC   -!- PATHWAY: Methylamine utilization.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: Contains 1 plastocyanin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57963; AAA68895.1; -.
DR   EMBL; L26406; AAB46937.1; -.
DR   PIR; A56621; CUPSAM.
DR   HSSP; P22364; 1AAC.
DR   InterPro; IPR000923; BlueCu_1.
DR   InterPro; IPR001235; Copper_blue.
DR   InterPro; IPR008972; Cupredoxin.
DR   Pfam; PF00127; copper-bind; 1.
DR   PRINTS; PR00156; COPPERBLUE.
DR   ProDom; PD001235; Copper_blue; 1.
DR   PROSITE; PS00196; COPPER_BLUE; 1.
KW   Copper; Electron transport; Periplasmic; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    119       AMICYANIN-ALPHA.
FT   DOMAIN       21    119       PLASTOCYANIN-LIKE.
FT   METAL        67     67       COPPER (BY SIMILARITY).
FT   METAL       106    106       COPPER (BY SIMILARITY).
FT   METAL       109    109       COPPER (BY SIMILARITY).
FT   METAL       112    112       COPPER (BY SIMILARITY).
SQ   SEQUENCE   119 AA;  12609 MW;  732FDECA8239D857 CRC64;

  Query Match             73.8%;  Score 31;  DB 1;  Length 119;
  Best Local Similarity   80.0%;  Pred. No. 52;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 CXPHP 7
              | |||
Db        106 CTPHP 110


RESULT 15
AMCY_PARDE
ID   AMCY_PARDE     STANDARD;      PRT;   131 AA.
AC   P22364;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Amicyanin precursor.
GN   MAUC OR AMI.
OS   Paracoccus denitrificans.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC   Rhodobacteraceae; Paracoccus.
OX   NCBI_TaxID=266;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCIMB 8944;
RX   MEDLINE=91085564; PubMed=2261991;
RA   van Spanning R.J.M., Wansell C.W., Reijnders W.N.M., Oltmann L.F.,
RA   Stouthamer A.M.;
RT   "Mutagenesis of the gene encoding amicyanin of Paracoccus
RT   denitrificans and the resultant effect on methylamine oxidation.";
RL   FEBS Lett. 275:217-220(1990).
RN   [2]
RP   SEQUENCE OF 27-36.
RX   MEDLINE=86243362; PubMed=3718960;
RA   Husain M., Davidson V.L.;
RT   "Properties of Paracoccus denitrificans amicyanin.";
RL   Biochemistry 25:2431-2436(1986).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY OF COMPLEX WITH MADH.
RX   MEDLINE=92287919; PubMed=1599920;
RA   Chen L., Durley R., Poliks B.J., Hamada K., Chen Z., Mathews F.S.,
RA   Davidson V.L., Satow Y., Huizinga E.G., Vellieux F.M.D., Hol W.G.J.;
RT   "Crystal structure of an electron-transfer complex between
RT   methylamine dehydrogenase and amicyanin.";
RL   Biochemistry 31:4959-4964(1992).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX   MEDLINE=94188715; PubMed=8140419;
RA   Chen L., Durley R., Mathews F.S., Davidson V.L.;
RT   "Structure of an electron transfer complex: methylamine
RT   dehydrogenase, amicyanin, and cytochrome c551i.";
RL   Science 264:86-90(1994).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.31 ANGSTROMS).
RA   Cunane L.M., Chen Z.-W., Durley R.C.E., Mathews F.S.;
RT   "X-ray structure of the cupredoxin amicyanin, from Paracoccus
RT   denitrificans, refined at 1.31-A resolution.";
RL   Acta Crystallogr. D 52:676-686(1996).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.3 ANGSTROMS).
RX   MEDLINE=99080123; PubMed=9860825;
RA   Zhu Z., Cunane L.M., Chen Z.-W., Durley R.C.E., Mathews F.S.,
RA   Davidson V.L.;
RT   "Molecular basis for interprotein complex-dependent effects on the
RT   redox properties of amicyanin.";
RL   Biochemistry 37:17128-17136(1998).
CC   -!- FUNCTION: Primary acceptor of electrons from methylamine
CC       dehydrogenase. Passes those electrons on either a soluble
CC       cytochrome c or to pseudoazurin.
CC   -!- COFACTOR: Binds 1 copper ion per molecule.
CC   -!- PATHWAY: Methylamine utilization.
CC   -!- SUBCELLULAR LOCATION: Periplasmic.
CC   -!- SIMILARITY: Contains 1 plastocyanin-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X55665; CAA39199.1; -.
DR   PIR; A24407; A24407.
DR   PIR; S12972; S12972.
DR   PDB; 1AAJ; 31-OCT-93.
DR   PDB; 1AAN; 31-OCT-93.
DR   PDB; 1MDA; 31-OCT-93.
DR   PDB; 2MTA; 31-JAN-94.
DR   PDB; 1AAC; 08-MAR-96.
DR   PDB; 1BXA; 07-OCT-98.
DR   PDB; 2RAC; 07-OCT-98.
DR   PDB; 1MG2; 11-DEC-02.
DR   InterPro; IPR000923; BlueCu_1.
DR   InterPro; IPR001235; Copper_blue.
DR   InterPro; IPR008972; Cupredoxin.
DR   Pfam; PF00127; copper-bind; 1.
DR   PRINTS; PR00156; COPPERBLUE.
DR   ProDom; PD001235; Copper_blue; 1.
DR   PROSITE; PS00196; COPPER_BLUE; 1.
KW   Copper; Electron transport; Periplasmic; Signal; 3D-structure.
FT   SIGNAL        1     26
FT   CHAIN        27    131       AMICYANIN.
```

```
FT   DOMAIN       27    131       PLASTOCYANIN-LIKE.
FT   METAL        79     79       COPPER (BY SIMILARITY).
FT   METAL       118    118       COPPER (BY SIMILARITY).
FT   METAL       121    121       COPPER (BY SIMILARITY).
FT   METAL       124    124       COPPER (BY SIMILARITY).
FT   STRAND       29     30
FT   STRAND       37     38
FT   HELIX        39     41
FT   TURN         44     45
FT   STRAND       47     52
FT   TURN         53     54
FT   STRAND       55     56
FT   STRAND       60     63
FT   TURN         65     66
FT   STRAND       68     73
FT   STRAND       79     79
FT   STRAND       82     83
FT   TURN         85     87
FT   STRAND       93     94
FT   STRAND       98     98
FT   TURN        100    101
FT   STRAND      102    108
FT   STRAND      112    118
FT   TURN        119    120
FT   STRAND      121    131
SQ   SEQUENCE   131 AA;  13983 MW;  F7352A865FD089DA CRC64;

  Query Match             73.8%;  Score 31;  DB 1;  Length 131;
  Best Local Similarity   80.0%;  Pred. No. 57;
  Matches    4;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 CXPHP 7
              | |||
Db        118 CTPHP 122


Search completed: April  8, 2004, 11:11:04
Job time : 5.83333 secs
```